AFIT/GSE/ENY/93D-2

CREATION OF PROTOTYPE AIRCREW PROTECTION EQUIPMENT BASED ON FACE ANTHROPOMETRY

THESIS

Martin E. Piccus, Captain, USAF
Gary A. Smith
Brian K. Standley, Captain, USAF
Tami L. Volk, Captain, USAF
Linda B. Wildes, Captain, USAF

AFIT/GSE/ENY/93D-2

AFIT/GSE/ENY/93D-2

CREATION OF PROTOTYPE AIRCREW PROTECTION EQUIPMENT BASED ON FACE ANTHROPOMETRY

THESIS

Presented to the Faculty of the Graduate School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment for the Degree of
Master of Science in Systems Engineering

Martin E. Piccus, Captain, USAF
Gary A. Smith
Brian K. Standley, Captain, USAF
Tami L. Volk, Captain, USAF
Linda B. Wildes, Captain, USAF

December 1993

Approved for public release; distribution unlimited

Preface

This research combines the disciplines of three dimensional data collection, spacial statistics, clustering, computer aided design and rapid prototyping into a sizing system for aircrew oxygen masks. It is the first successful attempt at merging these techniques for the purpose of producing representative faces for a sample population. This research is a distinct departure from the methods of the past. The final result provides a framework for creating better fitting masks and, as a consequence, could reduce the number of custom masks produced every year.

Special recognition is given to Major David G. Robinson. His vision and guidance was our impetus from conception to completion. It was his experience in the use of spacial statistics and his vision for its practical application which was the foundation for this research. We offer our collective thanks for all his patience and assistance in bringing this project to completion.

We would also like to thank the members of our committee: Lieutenant Colonel Paul Auclair, Captain Dennis Ruck, Captain Chris Hall, and Doctor Robert Beecher. Their expertise involving related disciplines for this research was invaluable.

We extend our gratitude to Mr Jay Anderson for his service and expertise in the use of the Stratasys Rapid Prototyper. We are also indebted to the men and women of the custom ask shop, Aerospace Physiology Directorate, 645th Medical Group, for their knowledge and assistance in the production of aircrew oxygen masks. Airman First Class John Shedio and Technical Sergeant Doug Holty deserve special acknowledgement for sharing their knowledge, time and energy in the creation of the prototype masks. We are further obliged to the Gentex Corporation for their hospitality in showing us how aircrew oxygen masks are currently produced.

This research was sponsored by the Life Support Systems Division, Human Systems Center at Brooks AFB, Texas. The support of Lieutenant Colonel Clemens has been much appreciated during this endeavor.

Last, but not least, we would like to thank our families for their patience and understanding during the past year-and-a-half. You made it all worthwhile.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AOM	Aircrew Oxygen Mask
BLUP	Best Linear Unbiased Predictor
CAD	Computer Aided Design
CMS	Custom Mask Shop
EDMA	Euclidean Distance Matrix Analysis
FESA	Finite Element Scaling Analysis
FDM	Form Difference Matrix
FM	Form Matrix
GLOC	G-Induced Loss of Conscienceness
ID	Identification Number
KLT	Karhunen-Loeve Transform
RPM	Rapid Prototyping Machine
SD	Spacial Delineators
SF	Safety Factor
TPS	Thin Plate Spline

Abstract

Emphasis on the human-to-aircraft interface has magnified in importance with the development of increased performance aircraft. The problem is fitting a largely variant aircrew population to a minimum number of different sized aircrew protection equipment. The focus of this research is the fit of positive pressure oxygen masks. While there have been significant advances in the design of aircrew oxygen masks, the sizing systems used for fitting these masks have not kept pace. Landmark-based sizing systems have consistently fallen short of fitting a large enough population, increasing the number of custom masks manufactured every year. The Life Support Systems Division, Human Systems Center, Brooks AFB TX is the sponsor of this research. The sponsor's interest in this research is based on their involvement in the design and manufacture of aircrew oxygen masks: the current contracted manufacturing process for standard sized masks (MBU-5/P and MBU-12/P); the custom mask shop located at Wright-Patterson AFB; and the future design and manufacture of the COMBAT EDGE mask, MBU-20/P.

The objective of this research is to develop a process which will analyze and group human faces,concentrating on the oxygen mask seal area, into a minimum number of generic yet distinct face types. The shape and size, or form, of the face is described by bending energies. This bending energy provides the features for clustering using fuzzy arithmetic. Fuzzy clustering then permits the grouping of similar face types, while distinguishing representatives for each cluster. The face tariffs are the percentage of faces grouped under a representative face. The representative faces are then transformed into three-dimensional solids using computer-aided design. A rapid prototyper creates the solids or plugs from which MBU-5/P masks are made and tested. The two major outcomes of this research are a formal morphological

method which allows the comparison between multiple faces, and a process to transform anthropomorphic surfaces into a prototype mask mold.

CREATION OF PROTOTYPE AIRCREW PROTECTION EQUIPMENT BASED ON FACE ANTHROPOMETRY

I. Introduction

An important element in maintaining air superiority is the human-to-aircraft interface. The research presented here is the first attempt to apply state-of-the-art data collection and reduction methods coupled with multivariate statistical analysis and fuzzy clustering techniques to determine face tariffs for oxygen mask designs. A face tariff is the percentage of faces grouped under a specific representative face type. The sponsor for this research is LtCol Clemens, Chief of the Life Support Systems Division, Human Systems Center, Brooks AFB TX.

1.1 Problem Statement

Current aircrew face masks, designed for a 1950's Caucasian male population with smaller cranial dimensions, fail to adequately fit the increasingly diverse population of aircrew members, as evidenced by growing demands for custom masks.

1.2 Background

Aircrew members depend heavily upon their oxygen mask. Unlike many aircraft components, there are no redundant systems for lungs. Mother Nature did not create electrical, mechanical, or hydraulic 'back-ups' to keep red blood cells enriched with oxygen in the event of primary failure. It is in this spirit that the aircrew member searches for that perfect fitting aircrew oxygen mask (AOM).

The two factors of historical significance making this quest more difficult are the ever-expanding flight envelopes for aircraft and the changing aircrew demograph-

ics. First, the modern jet fighter is capable of agility unthought of when jet aircraft were introduced at the close of World War II. Unfortunately, every increase in performance places an additional burden on the human operators, reaching the point where a need arises for the interjection of human-factors engineers and physiologists. These engineers design systems which attempt to prevent the onset of such maladies as spacial disorientation and G-induced loss of consciousness (GLOC) in high performance aircraft. Aircrews have also become keenly aware of the importance of a well fitting oxygen mask in this high-G environment and are increasingly selective in their choice of aircrew ensemble.

The second factor affecting the fit of current mask designs is the changing demographics among the flying community. When the United States Air Force (USAF) was created in 1947, the overwhelming majority of aircraft operators were white males. Forty-six years later, the flying community more accurately represents the diversification of the American people with the inclusion of minorities and women. Additionally, as Americans have been receiving better health care and nutrition, they have been growing larger, with a proportional increase in head size (39).

As the need for a properly fitting mask has increased, it has also become more difficult to find such a mask. The current operational solution for the most difficult fit problems is the 645th Medical Group, Aerospace Physiology Directorate, Custom Mask Shop or CMS, located at Wright-Patterson Air Force Base, Ohio. Creating custom masks is a time consuming and archaic operation, producing masks with an operating life of only 18 months. The change in demographics has increased the workload of the custom mask shop; need now exists for a new sizing plan.

1.3 User Needs

The effect of poorly fitting AOMs affects all levels of users, resulting in three identifiable needs. First, due to changing demographics and physiological evolution, the USAF needs an updated sizing system for determining future mask designs. The

second need is of the pilots and aircrews, who need better fitting masks to allow them to maintain control over their advanced, high-technology environment. Thirdly, the CMS needs standard size face plugs better representing the growing number of custom mask users (12), which would minimize the time creating custom masks. Unfortunately, the CMS is also hampered by manufacturing techniques developed more than 20 years ago for the MBU-5/P, the first positive pressure AOM developed by the United States Air Force. This research analyzes the custom mask user database for development of a sizing system because the CMS is located at Wright-Patterson AFB, and is a convenient source for facial data in the form of plaster face casts. The objective of this research is to meet all three of these needs. While the need for a custom mask shop may never go away, the efforts of this research will reduce their work load by providing the USAF with a better sizing system, and by providing the CMS with designs that better represent a larger percentage of the custom mask population. Finally, the aircrews will benefit by receiving better fitting masks.

1.4 Scope

The purpose of this research is to develop a process which analyzes and groups the human face, concentrating on the AOM seal area, into a minimum number of generic yet distinct face types. The outcome is the set of representative faces with corresponding face tariffs. A face tariff is the percentage of faces represented by each distinct face type. The representative faces are prototyped into three-dimensional (3-D) face 'plugs' from which MBU-5/P masks are made. This study is limited to the more difficult to fit face types: the growing number of aircrew members requiring custom made MBU-5/P masks. The end product is a process that compares the differences in form (shape and size) between faces, identifies the face tariffs, and translates face measurements into a prototyped plug. The accomplishment of this project also reduces the number of custom masks by producing standard plugs for

custom mask users. The scope is limited to identifying and producing face plugs, not modifying current face mask designs beyond the mask-to-face seal.

1.5 Research Objectives

The stated objective of this research is to develop a process which determines the minimum number and optimal form of standard size aircrew oxygen mask designs providing better fitting masks for a larger percentage of the aircrew population. This research is organized to solve the problems of establishing a better AOM sizing system, and reducing the workload for producing custom MBU-5/P masks. The research objectives break down this effort into separate pieces, establishing incremental goals, which when combined, provide solutions to the sizing system problems, addressing the users needs.

Collect/Prepare Data. The goal of this research objective is to collect as many samples/faces as possible and prepare them for analysis. From these samples, a protocol for measuring or digitizing the critical areas of the face necessary for determining form description is made. The measurements are then oriented to a common axis to ease comparison. Since the digitization is done by hand, measurement errors are reduced by processing data through a smoothing routine.

Develop Form Characterization. The purpose of developing an analytical method to describe the form of a face is to facilitate multiple face comparisons. Such an undertaking requires the review of existing anthromorphological methods and other form analytical methods. Several alternatives exist, but the decision to develop a new form characterization method is based on the type of data available and the ineffectiveness of past methods. The desired method should also be insensitive to measurement differences, and invariant to rotational or translational differences between faces.

Determine Characteristic Features. A feature is a term used to describe the element used for comparison in clustering analysis, as opposed to what might be normally considered a facial feature. Feature selection is the process of taking the 3-D form description and reducing the data to a set of unique features for each face. Two restrictions exist on the final feature set selected from each face. First, the feature sets must be the same size for input into the clustering routine. Secondly, the maximum number of features is dimensionally constrained by the sample size and the number of clusters. The key is identifying the optimum number of features for comparison. There exists some definable minimum and maximum number of features. The minimum describes the number of features below which there is insufficient data for clustering truly random data. More features provide more information which aid in better differentiating and grouping distinct faces. However, at some point more features provide unnecessary information and only confound the differences between individual faces. Due to dimensionality restrictions, more features can actually cause classification rates to decrease. The selection of critical features therefore characterizes the ability to define goodness of fit, where the critical features uniquely identify the form of a face for comparison.

Cluster Faces. A cluster of faces is a group of faces with identifiable similarities. The clustering algorithm distinguishes differences/similarities between faces and creates meaningful clusters. The number of clusters is an alterable, it can be changed at the discretion of the researchers, but is dependent on several factors. The minimum number of clusters determines the point at which distinct similarities between faces are captured. Below this minimum, faces less similar may end up being grouped together. The maximum number of clusters chosen is determined by cost. The cost is defined differently for various levels – obviously the highest cost is already being paid at the custom mask shop where each request receives a custom mask. At the Air Force level, the best way to ensure quality fitting mask designs would be to provide custom masks to all aircrew members, even

the 'normal' ones. Therefore, the maximum number of clusters is determined as a trade-off between the cost of providing each aircrew member their own custom mask, and achieving the best fitting mask per group of individuals. Once the clusters are determined, the next task is choosing and describing the face which best represents each group or cluster. This choice also affects the overall fit of the mask to members in the cluster.

Prototype Mask. The best way to determine the achievement of this research is to create a face mask for testing. First a face must be recreated from the data, then a plug is built which the CMS uses to create an MBU-5/P. In essence, the goal of this research objective is to create a prototype plug that the CMS can use, from materials compatible to both the rapid prototyper and the custom mask manufacturing process. The CMS cooperates in this effort by providing the raw data and building an MBU-5/P out of the prototype plug. This face mask is used for testing on the individuals from the cluster group.

Test Ability to Define Fit. The purpose of this research objective is to take the prototype plugs and evaluate how well they fit on the sample faces. This should provide information which verifies the applicability of the entire process to categorizing the human face.

1.6 Constraints

There are several constraining factors which limit the ability to optimally satisfy many of the research objectives. The greatest factors are the form of the source of data, and the different types and availability of hardware and software tools.

The data consists of plaster face casts. These face casts represent the unique set of hard-to-fit individuals. Therefore, the sample space does not include 'normal' faces. Some faces are also noticeably distorted due to the method of creating plaster face casts. Straws inserted in the mouth for breathing gave the face a sunken-in

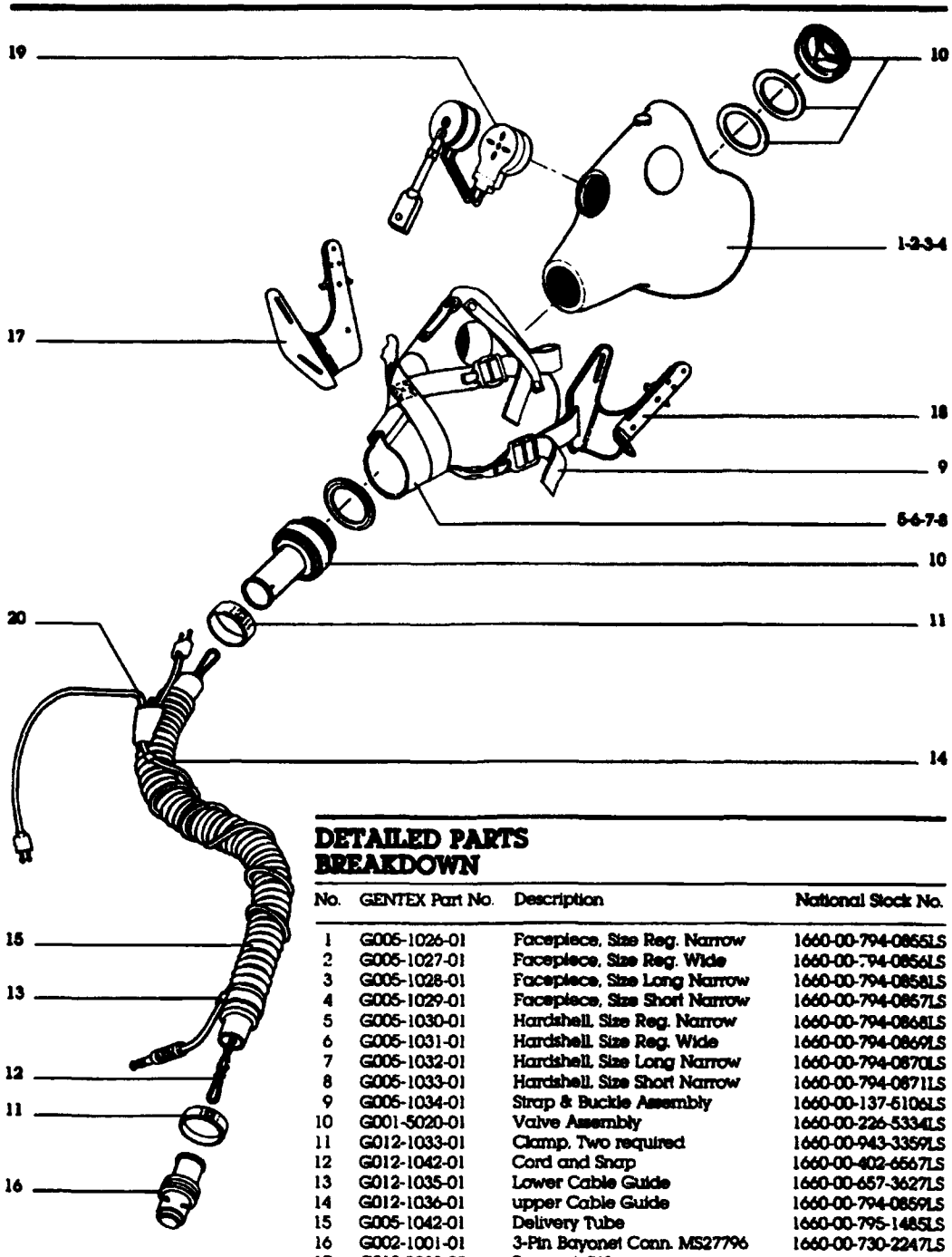
appearance, making the seal area narrower. Also, the plaster casts do not account for skin deformation under high-G loads. Some areas of the face may deform more easily than others, and these areas may require a 'tighter' fit. Finally, the choice of where to measure the faces could also affect the outcome of the clustering. This is dependent on the robustness of the form characterization method to inconsistencies in the data from face to face.

The hardware and software are chosen based on availability, ease of use, and learning curve. These three factors constrain development of the process. The hardware selection is based on availability and speed to produce usable results. The digitization requires a steady hand and including smoothing and orienting the data, is labor intensive. The software is developed to avoid spending excessive time programming. Therefore, most of the data reduction including feature selection is accomplished using MATHEMATICA routines. The clustering algorithm, which is written in FORTRAN, is the only stand-alone code written for this project.

One outcome of this project is a process for categorizing faces into a minimum number of face types. Once completed, this process might need to be periodically re-accomplished, due to changing aircrew demographics and anthropometry. Consequently, the AOM database needs to be updated and 're-clustered' at appropriate intervals.

1.7 Data Source

This research uses the custom mask user population as the source of facial data in the form of plaster face casts, since the CMS is conveniently located at Wright-Patterson AFB. However, it is important to realize that the custom mask users represent the unique set of aircrew members, whose faces do not fit the standard MBU-5/P masks. Their request for custom masks could stem from a variety of reasons, from ethnicity or sex (recall the current mask designs are based on a study involving only Caucasian males) to environmental factors (i.e., broken noses).



DETAILED PARTS BREAKDOWN

No.	GENTEX Part No.	Description	National Stock No.
1	G005-1026-01	Facepiece, Size Reg. Narrow	1660-00-794-0855LS
2	G005-1027-01	Facepiece, Size Reg. Wide	1660-00-794-0856LS
3	G005-1028-01	Facepiece, Size Long Narrow	1660-00-794-0858LS
4	G005-1029-01	Facepiece, Size Short Narrow	1660-00-794-0857LS
5	G005-1030-01	Hardshell, Size Reg. Narrow	1660-00-794-0868LS
6	G005-1031-01	Hardshell, Size Reg. Wide	1660-00-794-0869LS
7	G005-1032-01	Hardshell, Size Long Narrow	1660-00-794-0870LS
8	G005-1033-01	Hardshell, Size Short Narrow	1660-00-794-0871LS
9	G005-1034-01	Strap & Buckle Assembly	1660-00-137-5106LS
10	G001-5020-01	Valve Assembly	1660-00-226-5334LS
11	G012-1033-01	Clamp, Two required	1660-00-943-3359LS
12	G012-1042-01	Cord and Snap	1660-00-402-6567LS
13	G012-1035-01	Lower Cable Guide	1660-00-657-3627LS
14	G012-1036-01	upper Cable Guide	1660-00-794-0859LS
15	G005-1042-01	Delivery Tube	1660-00-795-1485LS
16	G002-1001-01	3-Pin Bayonet Conn. MS27796	1660-00-730-2247LS
17	G012-9030-02	Bayonet, RH	1660-00-186-0277LS
18	G012-9030-01	Bayonet, LH	1660-00-186-0276LS
19	G005-9002-01	Dynamic Microphone M100/AIC	
20	G012-9026-01	Communication Cord & Plug	5995-00-818-5368

Figure 1.1 MBU-5/P Mask Parts: Part 1 is the faceform, and part 5 is the hardshell (25).

The CMS currently makes 300 - 400 custom MBU-5/P masks per year. The process is very long and manpower intensive. The first step in making a custom mask is for the user to send a plaster face cast to the lab. This cast is created at the local hospital/dental clinic and represents a 3-D positive image of the individual's face. Some casts are omitted from the study due to face deformations during the process of creating the plaster face casts.

The CMS technicians draw a line indicating the mask seal on the plaster face cast. The line is used as a guide in the construction of a face plug which mates to the inside of the hard-shell (see Figure 1.1). The plug is made from hydrocal or dental stone, and after careful reshaping via scraping and sanding away excess material, the plug is coated with two layers of an epoxy finish which protects the surface and shape of the plug. This plug is dipped in latex, which after proper curing procedures, is carefully peeled off and affixed to the hard-shell creating the final mask. The most labor intensive part of this procedure is making the plug. The technicians spend an extra 20 - 30 minutes to test a plaster face against existing plugs (stored six years) to try and find a match, and avoid making another plug. If the incoming face matches an existing plug, the face is classified as Category V. If a face plug has to be made the incoming face is classified as Category I. Categories II, III, and IV are replacements for the faceform, hard-shell, and both faceform and hard-shell respectively, and do not require creation of a face plug.

1.8 Technical Approach

The purpose of this research effort is to statistically analyze anthropometric data to support improvements in the design of aircrew oxygen masks (AOMs) and use this information to create prototype representative custom masks. To accomplish these goals, the subject face was defined in terms of its bending energy; that is, the energy required to deform a flat plane to match the surface geometry of the face. This energy is analyzed to determine several spacial features, like nose height, chin

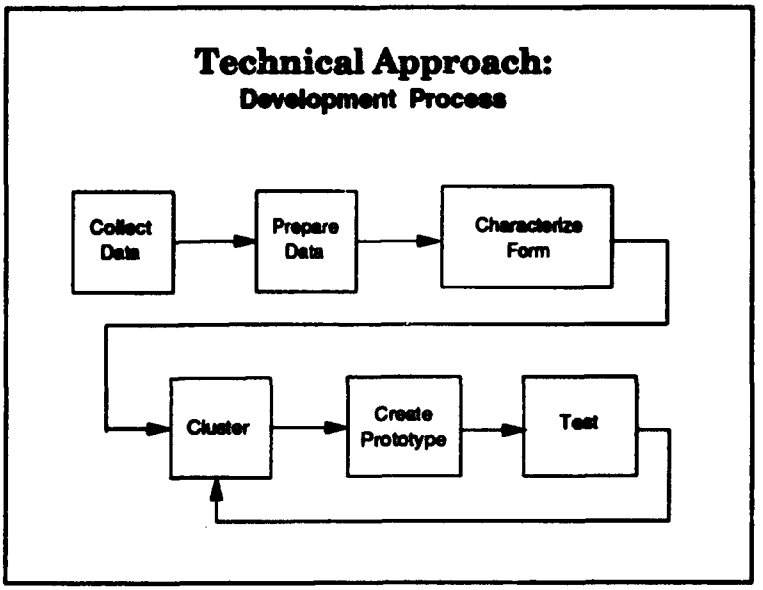


Figure 1.2 Technical Approach: Development process.

height, etc., of a human face. These features are then clustered and the resulting groups are analyzed to determine standard aircrew mask designs. The final designs are then created and tested. Figure 1.2 depicts the sequence of steps. A description of each step in the process is summarized below. Further details are provided in subsequent chapters.

Data Collection. Facial data is obtained from the plaster face casts collected by the CMS. The mask seal line, line drawn by lab technicians, and spacial delineators are used as the crucial measurements in the differentiation of face types. A 3-D digitizer is used in continuous mode to obtain (x, y, z) coordinates of the seal line. Additional points, providing spacial delineation within the seal area, are taken as individual points called spacial delineators to better capture facial features.

Data Preparation. Once the seal line and spacial delineator data (together they define the seal area) are obtained, each file is oriented against a

similar axis to facilitate input into the computer aided design (CAD) program called Pro/ENGINEER (41). To accomplish this, an origin for each seal area is calculated using center of mass calculations on the seal area data. Principal component analysis is used to orient the seal area. Pro/ENGINEER is also used to check the seal line for gross errors. These errors are removed and the line is then 'smoothed' transcribing the rough, hand digitized curve to a smooth curve.

Characterize Form. The seal area data is statistically analyzed via the form characterization method. This method provides a way for accurately describing the form of each face and the features necessary for comparison.

Cluster. The faces are clustered using the features from the form description. The cluster analysis produces groups of similar shapes. From these groups, a number of tariffs are defined, and a representative for each tariff is determined.

Prototype. The seal area data for each cluster or tariff representative is transferred to Pro-Engineer and mated with an outer shell of an MBU-5/P mask – thus creating a solid image. Once this is accomplished, the data is transferred to a prototyping machine, a Stratasys 3D MODELER, and a plug is created from plastic or wax. The material of the prototyped plug is determined by its compatibility with the custom mask manufacturing process.

Test. Testing is done in a number of ways at a number of different stages during the project. The tests include repeatability of digitizing the seal areas, ensuring Category V's which use the same plug are grouped together, and testing masks made from the prototype plugs.

1.9 Summary

The basis for this research is the identification of a problem requiring an analytical solution approach. The ability to cluster and model face shapes is crucial towards designing a robust face mask. On a smaller scale, the MBU-5/P mask is the only mask that is custom manufactured for users that have unique or hard-to-fit faces. By digitizing the seal areas, the faces can be placed in a form that is easily manipulated and studied. The unique capability provided by statistical clustering using multivariate spacial statistics allows comparisons to be made of the complex and diverse data. By endeavoring to assign faces to categories or face types, this project effectively develops a process to cluster 3-D shapes into distinct groupings that will enable a few masks to better fit a larger population of individuals. The CAD and prototyping capability puts the reduced data into a piece of hardware for testing and custom mask manufacture. The end product enables human factors design considerations to impact future aircrew protection equipment in a way that will significantly improve the ability to successfully accomplish the overall mission.

The following chapters explain how these objectives are accomplished. Chapter II summarizes previous research in this area. Chapters III and IV develop the process for characterizing and grouping human faces, and Chapter V explains the methodology applied to the data and the results.

II. Literature Review

This chapter provides a review of the literature pertaining to two main conceptual areas of the present research. These areas are anthropometric studies of Air Force personnel and current methods of form description.

The review of anthropometric studies focuses on the work done since 1950 to measure and classify the craniofacial geometries of aircrew members. These studies led directly to the sizes of aircrew oxygen masks currently used by the USAF, and thus are at the root of present-day fit deficiencies. In the years since these studies were carried out, new methods for analyzing geometric data that express shape and size have been developed.

In exploring techniques for describing and analyzing form and form differences, this chapter also examines several methods used in the field of morphometrics. These methods are procrustes, finite element scaling analysis, Euclidean distance matrix analysis and thin plate splines. The advantages and limitations of each method are identified, and conclusions are reached on their usefulness in the present study.

2.1 Anthropometric Studies

The USAF has conducted several studies to collect physical measurements of its personnel, and determine standard sizes of clothing and equipment. This section reviews past and current efforts in the characterization of face shapes for the sizing of aircrew oxygen masks (AOM).

2.1.1 1950 Survey. In 1950, a survey of anatomical measurements was conducted by the USAF Anthropometric Laboratory, located at Wright-Patterson AFB, in conjunction with Antioch College (28). The objective of the study was to gather a large data base to be used in the design of cockpits, flying equipment, and apparel. The data were gathered by a survey team that traveled to 12 Air Force bases

around the country to take physical measurements of individual aircrew members. This 1950's study measured 132 body features, including 41 from the face alone. The tools used to conduct the study consisted of calipers, rulers, and tape measures. The survey team consisted of 12 students from Antioch College who went through extensive training to ensure the measurements taken by any one student could be duplicated by any other. The goal of the study was to collect measurements and demographic data from 5,000 aircrew members, but the outbreak of the Korean War cut it short, with only 4,000 individuals measured.

The information gathered in the survey was analyzed and categorized. By 1957, the 41 head measurements had been cataloged according to relationships between different facial features. Correlation coefficients were calculated in an attempt to determine which face measurements were strongly related and which were not (9). The results of this statistical analysis led to the conclusion that the statistical relationships between head and face dimensions are quite low, and that procedures used in sizing equipment for use on other parts of the body; that is, selecting sizing dimensions so that variability of other dimensions is substantially reduced; could not be applied in this instance (22). Based on previous experience, the designers selected overall face length and width as the critical dimensions, and decided upon six standard sizes for the prototype MBU-5/P oral-nasal oxygen mask. The sizes were designated short, medium, and long, with each of these having a narrow and wide version. Other mask dimensions were fixed by taking the mean measurement in each size group and adding or subtracting a standard deviation, depending on the particular dimension. For example, the width of the mask in the nasal bridge area was adjusted downward from the mean to insure a tight fit, to prevent leakage, and reduce obstruction of the wearer's vision. The portion of the mask that fit over the rest of the nose, however, was sized larger than the mean to avoid pinching in this area where a snug fit is less critical. The percentage of personnel who would be wearing a particular size, or tariffs, were also obtained from the statistical parame-

ters computed in the survey. A fit test was conducted to confirm the validity of the six sizes. This test found that the short-wide and the long-wide sizes were unnecessary, since the rubber face seal of the short-narrow and the long-narrow were pliable enough to compensate for the greater width of an individual's face. The medium-wide was retained because the testing indicated a relatively large percentage of the population would require this size.

2.1.2 1967 Study. In 1967, another survey of anthropometric data was conducted to support the design of flight equipment (56). There were 2,420 rated male officers from 17 bases who had 187 body dimensions measured, of which 48 were on the head and face. The study found that in the 17 years since the original survey, head circumference had increased by 4.9mm, face length had increased by 2.7mm, and face breadth had increased by 1.4mm, on average. These new data were used in the same regression equations utilized in 1957, and again very little correlation was found between any pair of dimensions.

These data were used in the design of a new mask to replace the MBU-5/P. This time, face length alone was selected as the size discriminator, based on the finding that, for the MBU-5/P, a choice of widths was unnecessary (39). Mean and standard deviation values for each dimension were calculated within each of the size groups, and these were applied in designing the overall mask geometry. This new mask, designated MBU-12/P, was designed for higher G-loads and the incorporation of chemical warfare gear. During the fit test and evaluation of the MBU-12/P in 1975, the overwhelming majority of aircrews participating in the study preferred the new mask to the MBU-5/P in the areas of ability to valsalva, ability to maintain pressure at higher altitudes (greater than 14,000 meters), lateral head mobility, and slippage during increased G-loads (13).

2.1.3 Current Focus. Since the 1970s, it has become apparent the MBU-12/P and traditional G-suit would not be adequate for the future and, to some extent,

current aircraft. The COMBAT EDGE system was designed to fight the potential for G-induced loss of consciousness (GLOC) in high-G maneuvers. COMBAT EDGE is composed of a modified HGU-55P helmet, a CSU-138/P anti-G suit, an external counterpressure vest, and a MBU-20/P positive pressure oxygen mask. Studies have been conducted on the entire ensemble. Although the MBU-20/P was said to ease breathing during high-G flight profiles, a drawback cited was "distraction due to mask leakage" (38). Mask leakage was typically found to occur around the nose bridge and under the eyes, where the escaping high pressure oxygen dried the eyes and interfered with vision. Clearly, quality of fit is still a critical factor in the design of oxygen masks.

2.1.4 Summary. The analytical techniques used previously to determine mask shapes and sizes relied on the tools and statistical methods of those times. This will be the first Air Force research project to use digitized face geometry data and to employ computer aided design (CAD) technology in processing that data. It will also be the first to use modern form analysis techniques and fuzzy clustering to classify face shapes.

2.2 3-Dimensional Form Description

This section covers four existing form analytical methods (also referred to as morphological methods), all of which have been used for analyzing anthropometric differences or similarities. The four methods discussed are: procrustes, finite element scaling analysis (FESA), Euclidean distance matrix analysis (EDMA) and thin plate splines (TPSs). The next few paragraphs describe some basics pertinent to form analysis. Then, each of the four morphometrical methods researched are explained. The explanations include how each method is used in form analysis, assumptions made, and advantages and disadvantages of using the method. A summary covers the applicability of each method to this research, and identifies the need to develop a new form characterization method.

Form. The 'form' of an object refers to both size and shape. Form is represented by data collected from the object. The form of an object is defined to be that characteristic which remains invariant under translation, rotation, and reflection of the object. Shape on the other hand is defined to be the characteristic which remains invariant under translation, rotation, reflection and scaling of the object (35).

Morphometrics. Morphometrics is the study of form analysis. It has been described as "the empirical fusion of geometry with biology" (3). "A true merger of geometry and biology requires that a biological object can be unambiguously reconstructed from the data collected to represent that particular biological object" (46). Past U.S. Air Force morphometric analyses, such as the multivariate approaches conducted in the 1950's, could not reconstruct the biological object from the linear distances collected. The linear distances measured between landmarks were thought to be points which best distinguished characteristics of the biological object. Past analyses usually resulted in a series of traits that described particular features but did not define the geometry of the object (46).

Landmarks. Landmarks must be consistently present on all objects under consideration to be useful in analysis. The minimum criterion for landmarks is that a landmark must be "consistently and reliably located with a measurable degree of accuracy on all objects considered" (36). It is important to note that landmarks alone do not provide all information pertaining to the form of an object. When using landmarks in morphometric analyses, information on the curvature and other surface features between landmarks is lost (46).

Coordinate Dependency. There are two general classes of morphometric method analysis: coordinate-based and coordinate-free. In a coordinate-based method,

> ... results can be rotated to any coordinate system without change in or loss of information, and a coordinate system is a necessary part of the analytical machinery. Coordinate-based methods measure form difference as a deformation from a reference to a target form, or as the fit resulting from the superimposition of two forms (46).

In the coordinate-based method, any arbitrary coordinate system can be chosen. On the other hand, coordinate-free methods analyze form differences without reference to any coordinate system. This can be accomplished by using distances between landmarks as in EDMA. There is a very subtle difference between the two classes which does not become apparent until comparisons between objects are made. It is particularly distinguishable in comparisons of patterns of form-change through time, such as growth analyses of biological objects.

2.2.1 Procrustes. Procrustes is a method which attempts to quantify the differences between the forms (shapes and sizes) of two objects (35). It is named after the mythological Greek character who would tie all travellers that crossed his path to an iron bed. If the traveller was shorter than the bed, Procrustes would stretch his limbs until he fit it. If the traveller was longer than the bed, Procrustes would hack off whatever bits hung over the end (7).

The method of procrustes requires that the two forms being compared have a one-to-one correspondence of landmarks. The landmarks must be points that can be clearly and distinctly defined on both the reference form and the target form; that is, there must be a one-to-one correlation between the landmarks on the two objects. In the method of procrustes, one places an object directly over another object and calculates the squared distances between corresponding points. These squared distances are summed and the resulting number is a measure of how different one object (the target object) is from some reference object. In equation form, this

summation is:

$$\sum_{i=1}^{n} d^2(L_i, A_i) = \text{measure of similarity} \tag{2.1}$$

Where:

$$d^2(L_i, A_i) = (x_{L_i} - x_{A_i})^2 + (y_{L_i} - y_{A_i})^2 + (z_{L_i} - z_{A_i})^2 \tag{2.2}$$

and L_i is point i on the reference object, A_i is point i on the object to be transformed, and n is the number of points on each object. In matrix form, this equation is equivalent to calculating the trace of:

$$\begin{bmatrix} d^2(L_1, A_1) & 0 & 0 & \dots & 0 \\ 0 & d^2(L_2, A_2) & 0 & \dots & 0 \\ 0 & 0 & d^2(L_3, A_3) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d^2(L_n, A_n) \end{bmatrix}$$

The procrustes optimal position is obtained by rotating and translating the target object until a minimum sum of squared distances between corresponding landmarks of the target and reference object has been achieved.

2.2.1.1 Procrustes Example. A simple example of procrustes analysis follows. Figure 2.1 shows four simple objects. Object L is the reference object and objects A, B, and C are the objects to be transformed. Figure 2.2 is object A placed directly over object L. The positioning of one object over another can take on different forms. In ordinary procrustes, they are positioned such that each point on the transformed object is equidistant from the corresponding point on the reference object. The squared Euclidean distances between corresponding points are calculated, $d^2(L_i, A_i) = 1.25$, $i = 1, 4$ and these distances are summed together,

$\sum_{i=1}^{n} d^2(L_i, A_i) = 5.0$. The number 5.0 is an indication of how different object A is from object L.

2.2.1.2 Disadvantages. Procrustes analysis has some inherent weaknesses. The first shortcoming is in the meaning of the final answer, 5.0. This solution is not unique between objects L and A. Figure 2.3 is the procrustes comparison of objects B and C with object L. Both these objects also yield a final solution of 5.0 and yet they are not the same size and shape as object A. Analysis of this shortcoming brings to light other problems. Once a sum of squared distances between points has been calculated, there is no method for recreating the original figure using this single scalar value. This value is an indication of a global shape change between two objects and gives no indication of local variations in shape. Lastly, to use procrustes, obviously there needs to be the same number of landmark points on both objects being compared.

In the field of morphometrics, the objective is to transform the biological form of one entity into another similar entity and measure this sum of squared distances. For the task of characterizing face shapes, the challenge is to calculate distances between identifiable and representative landmarks which give an indication of how alike two faces are. The problem is identifying landmarks. On the human face, there are not many clearly identifiable points which can be matched without palpation of the skin between subjects. The face is composed of many intersecting curves which have a high variability between subjects. Even if it was feasible to identify definitive landmarks among all humans, it is possible for two faces to have completely different appearances as judged by the human eye and have the same procrustes measure, as seen in the previous example. Local facial variations may be equally or more important than global variations. Procrustes analysis does not reveal any information about local variations. For these reasons, lack of clear, consistent landmarks and the lack of an indication of local variations, procrustes analysis is inappropriate for this research.

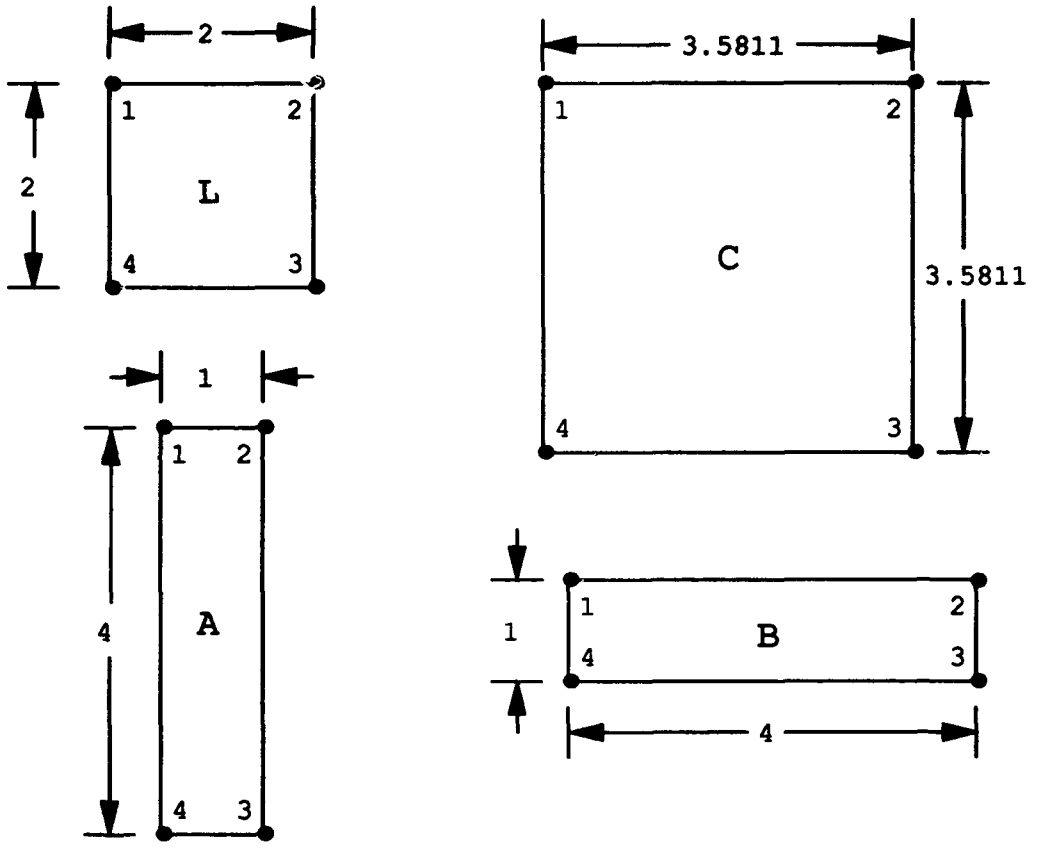


Figure 2.1 Procrustes Example Objects.

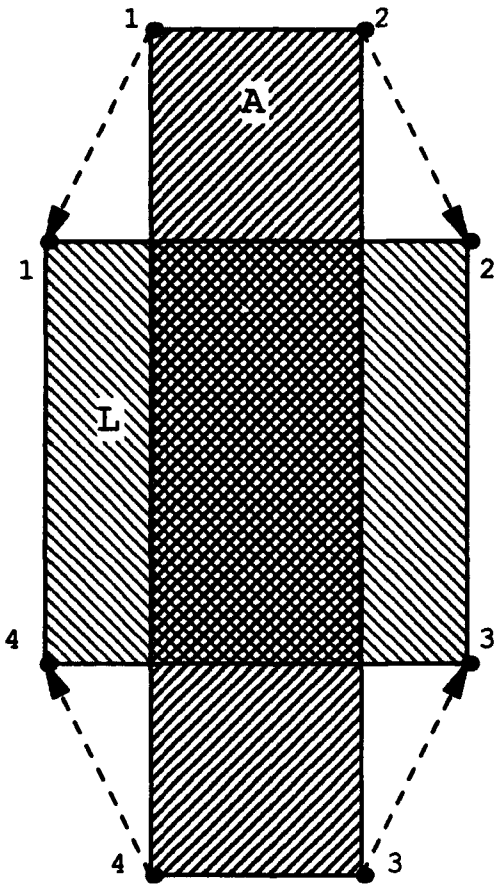
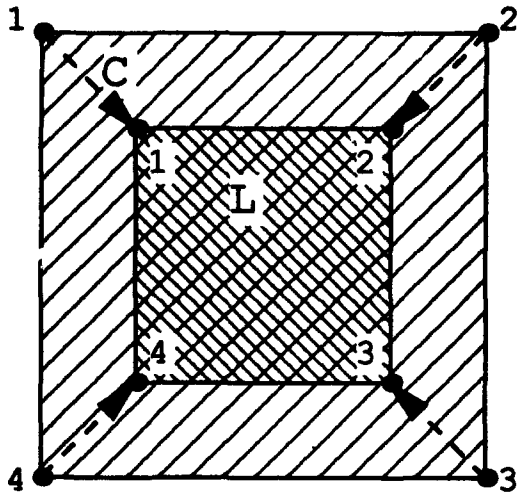


Figure 2.2 Procrustes Example 1: Transforming object A into reference object L.



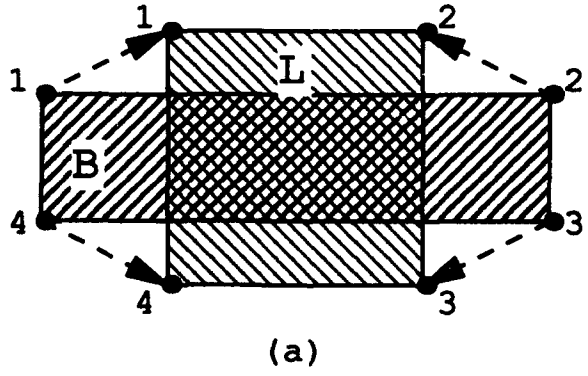


Figure 2.3 Procrustes Example 2: (a) Transforming object B into reference object L (b) Transforming object C into reference object L.

2.2.2 Finite Element Scaling Analysis. Finite element scaling analysis (FESA) is another method of comparing the geometries of two physical forms. It is closely related to the finite element analysis used in structural engineering to predict the deformation of a structure after the application of a set of forces (46). In FESA, a comparison is made between a reference form consisting of a set of landmarks on an object, and a target form consisting of the same set of landmarks on a second object.

2.2.2.1 General Description. As with the procrustes method, the landmarks must be points that can be clearly and distinctly defined on both the reference form and the target form; that is, there must be a one-to-one correspondence between the landmarks on the two objects. The landmarks are connected by line segments to make a series of finite elements that together make up the model under study. In three-dimensions, these elements typically are four-noded tetrahedrons or eight-noded cubes. The deformation, or morphological strain, required to produce the target form from the reference form is calculated element by element and expressed as a 'form tensor'. Note that in FESA, the term 'strain' simply means the difference in form between two objects, and is not a result of the application of external forces.

The form tensor is a symmetric 3-by-3 matrix (when dealing with three-dimensional forms) containing all information about morphological change occurring local to each landmark considered. Orthogonal principal directions, or eigenvectors, can be extracted from the form tensor, and these identify the directions of maximum, intermediate, and minimum difference between the two forms. The eigenvalue associated with each principal direction gives the magnitude of the deformation along that axis. Thus, six values (three angles and three linear distances) are required to completely describe the strain at each landmark. This information can be used in any number of ways depending on the aim of the particular study. For example, the change in distance between each pair of landmarks may be calculated.

2.2.2.2 Advantages. FESA is a coordinate-based method in that the two forms being compared must have their landmark positions measured in the same global coordinate system. But the method has the advantage of being registration-free, meaning that each form can have any orientation within that coordinate system, with no regard for how its landmarks align with those of the other form. Translation and rotation in going from the reference to the target form are eliminated. Deformations are measured independent of any specific coordinate system and the movement of the landmarks is not expressed relative to a specified center (45). Rather, differences are measured local to each landmark, and can be rotated into any coordinate system desired. This gives FESA the valuable ability to localize morphological differences between objects to the areas around particular landmarks, and to identify the anatomical dimensions along which two objects differ the most and the least.

2.2.2.3 Disadvantages. As with procrustes, most studies involve comparing multiple target forms, one at a time, with the reference form. The intra-landmark dimension that shows the greatest difference between the reference form and a particular target form is the best descriminator between them. However, that dimension is not necessarily the best descriminator between the reference form and any other target form in the group under study. This requires the selection of landmarks defining anatomical dimensions that are best for a population as a whole, and therefore not necessarily optimum for any individual target form. The larger the population under study, the more difficult it becomes to select landmarks that adequately distinguish between all of the forms.

The dependence of FESA on landmarks makes it unsuitable for this research. As mentioned previously, the geometry under consideration, the human face, is not easily described by landmarks. The line of contact between the oxygen mask seal and the wearer's face is a complex, three-dimensional curve. It varies widely from face to face, and the selection of landmarks that adequately capture the geometry of each face is impossible.

2.2.3 Euclidean Distance Matrix Analysis. This section gives a general description of Euclidean distance matrix analysis (EDMA). Discussion includes development of the two basic matrices used in EDMA, the form matrix and form difference matrix. Examples are added to demonstrate how EDMA is applied to compare objects. The major advantages and disadvantages of this method are highlighted in the summary.

2.2.3.1 General. EDMA, like procrustes and finite element scaling analysis (FESA) described in previous sections, is also based upon landmark data and also uses matrices to track information. EDMA is a coordinate-free morphological analysis technique and being coordinate-free, this method is indifferent to rotation, translation and reflection. EDMA can be used for two-dimensional or three-dimensional data. Discussion on EDMA covers the more difficult three-dimensional applications: however, for simplification, three two-dimensional examples are given. Given k landmarks characterized by (x, y, z) coordinates in three-dimensional space, an object A can be represented by a $k \times 3$ matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \\ \vdots & \vdots & \vdots \\ x_k & y_k & z_k \end{bmatrix}$$

The information contained in this matrix is used to calculate the form matrix, a basic matrix calculated for every object considered in EDMA analysis.

2.2.3.2 Form Matrix. Given a matrix $\mathbf{A}$ of k landmarks, the Euclidean distance matrix, or form matrix (FM), can be calculated. The Euclidean distance matrix is a symmetric $k \times k$ matrix where each element (i, j) corresponds to the Euclidean distance between landmarks i and j. This matrix contains information

on both size and shape of the object and is mathematically designated by **F**. The FM is a function of the matrix consisting of landmark data, **A**, which represents object A :

$$\mathbf{F(A)} = \begin{bmatrix} 0 & d(1,2) & d(1,3) & \cdots & d(1,k) \\ d(2,1) & 0 & d(2,3) & \cdots & d(2,k) \\ d(3,1) & d(3,2) & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & d(k-1,k) \\ d(k,1) & d(k,2) & \cdots & d(k,k-1) & 0 \end{bmatrix} = [F_{ij}(\mathbf{A})] \quad (2.3)$$

where $i = 1,\ldots,k$; $j = 1,\ldots,k$; and $d(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ (Euclidean distance between points i and j)(36).

2.2.3.3 Comparison of Two Objects. To compare one object to another, the two objects must have the same number of landmarks, and the landmarks must be in the same order for comparison; this means that in the $k \times 3$ coordinate matrix **A**, a landmark on object A must have a corresponding landmark on object B. Of course the (x, y, z) data will most likely be different. For instance, the tip of the nose on one person must be compared to the tip of the nose on another. Another matrix, the form difference matrix (FDM), is created to allow the comparison between two objects. The FDM, **D**, is a function of two matrices, **A** and **B**, where **A** and **B** represent the landmarks of objects A and B: **D(A,B)**. When comparing two objects, one is designated as the reference. For discussion purposes the reference object is A, and B represents the target object used for comparison. The off diagonal elements of the form difference matrix **D(A,B)** are generated by calculating a ratio of distance between two landmarks in object A with the corresponding distance in object B. The terms along the diagonal of the FDM are defined to be zero. The $(i,j)^{\text{th}}$ element of the form difference matrix is a ratio of the $(i,j)^{\text{th}}$ elements of the

form matrices of object A to object B when $i \neq j$:

$$\mathbf{D(A,B)} = [D_{ij}(\mathbf{A},\mathbf{B})] = [F_{ij}(\mathbf{A})/F_{ij}(\mathbf{B})] \text{ when } i \neq j$$

$$\mathbf{D(A,B)} = [D_{ij}(\mathbf{A},\mathbf{B})] = 0 \text{ when } i = j$$

$$\mathbf{D(A,B)} = \begin{bmatrix} 0 & \frac{F_{12}(A)}{F_{12}(B)} & \frac{F_{13}(A)}{F_{13}(B)} & \cdots & \frac{F_{1k}(A)}{F_{1k}(B)} \\ \frac{F_{21}(A)}{F_{21}(B)} & 0 & \frac{F_{23}(A)}{F_{23}(B)} & \cdots & \frac{F_{2k}(A)}{F_{2k}(B)} \\ \frac{F_{31}(A)}{F_{31}(B)} & \frac{F_{32}(A)}{F_{32}(B)} & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \frac{F_{k-1,k}(A)}{F_{k-1,k}(B)} \\ \frac{F_{k1}(A)}{F_{k1}(B)} & \frac{F_{k2}(A)}{F_{k2}(B)} & \cdots & \frac{F_{k,k-1}(A)}{F_{k,k-1}(B)} & 0 \end{bmatrix} \tag{2.4}$$

The form difference matrix is a symmetric matrix. When all off-diagonal elements are equal to one, the two objects have the same form, i.e. the landmarks of objects A and B have identical locations. This is not readily identifiable directly from the individual (x, y, z) coordinate data because the objects may have different coordinate systems. If all off diagonal elements are equal to some constant (other than one), then objects A and B have the same overall shape, but differ in size (scaling of one object). Typically, the off diagonal elements of the FDM are different. Their values indicate whether the Euclidean distances between the landmarks on object B are greater than the corresponding distances for reference object A (element D_{ij} would be less than one) or less than the corresponding distances for reference object A (element D_{ij} would be greater than one). By calculating the FDM and comparing all possible linear distances, EDMA provides a means for identifying 'local' areas of form difference (14). Only local areas can be evaluated since the analysis is limited to landmark information. No information regarding the areas or curvatures between landmarks on either object can be obtained. To illustrate the methodology a number of two-dimensional examples follow.

EDMA Example 1: Two objects of the same form. This example shows how EDMA easily detects two objects of the same form. Two squares A and

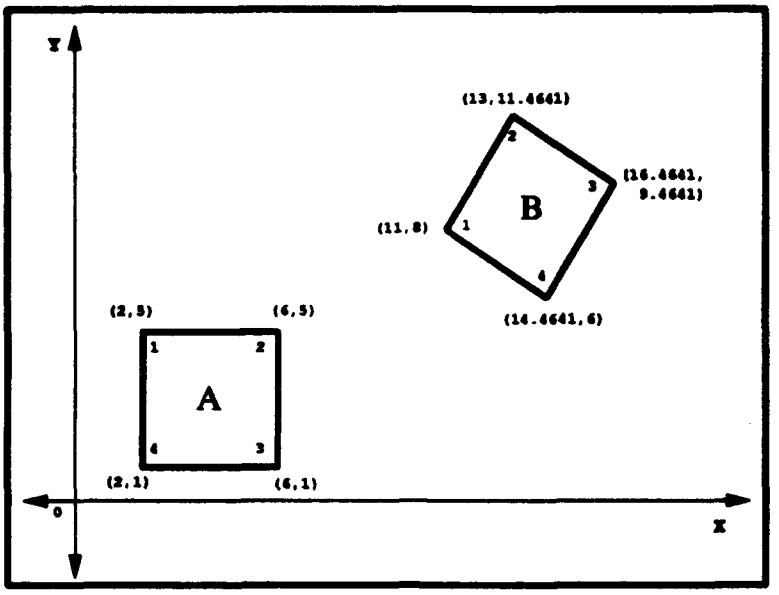


Figure 2.4 EDMA Example 1: Two objects of the same form.

B are given in Figure 2.4. They are the same size and shape, but have different coordinates. The corners of the squares represent landmarks. Objects A and B could come from two different coordinate systems, but are shown together for convenience. The matrices of the (x, y) coordinates of the four landmarks for A and B in Figure 2.4 are shown below. Note that the first row of each matrix contains the coordinates of landmark 1; the second row, landmark 2; etc.:

$$\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 6 & 5 \\ 6 & 1 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 11 & 8 \\ 13 & 11.4641 \\ 16.4641 & 9.4641 \\ 14.4641 & 6 \end{bmatrix}$$

The next step is creating the form matrix for each object. The distances between every pair of landmarks on each object are calculated using Equation 2.3. For example, the distance between landmark 1 and 3 for object A is $\sqrt{(2-6)^2 + (5-1)^2} = 5.6569$;

this is element $F_{13}(\mathbf{A})$. The FMs of objects A and B are:

$$\mathbf{F(A)} = \begin{bmatrix} 0 & 4 & 5.6569 & 4 \\ 4 & 0 & 4 & 5.6569 \\ 5.6569 & 4 & 0 & 4 \\ 4 & 5.6569 & 4 & 0 \end{bmatrix}$$

$$\mathbf{F(B)} = \begin{bmatrix} 0 & 4 & 5.6569 & 4 \\ 4 & 0 & 4 & 5.6569 \\ 5.6569 & 4 & 0 & 4 \\ 4 & 5.6569 & 4 & 0 \end{bmatrix}$$

In order to compare objects A and B, the form difference matrix must be calculated by using Equation 2.4. Using **A** as the reference matrix, the $(i,j)^{\text{th}}$ element of the FDM is found by creating ratios of the $(i,j)^{\text{th}}$ elements of the form matrices calculated above. For example, $D_{13}(\mathbf{A},\mathbf{B}) = \frac{F_{13}(\mathbf{A})}{F_{13}(\mathbf{B})} = \frac{5.6569}{5.6569}$. The FDM of objects A and B is:

$$\mathbf{D(A,B)} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

The off-diagonal elements of the FDM equal one. This illustrates that A and B do indeed have the same form (shape and size), regardless of any rotation, translation, and reflection which might be present.

EDMA Example 2: Two objects of the same shape. This example demonstrates how EDMA is able to capture the shape similarity and the size difference of new objects A and B. Figure 2.5 shows two squares A and B which

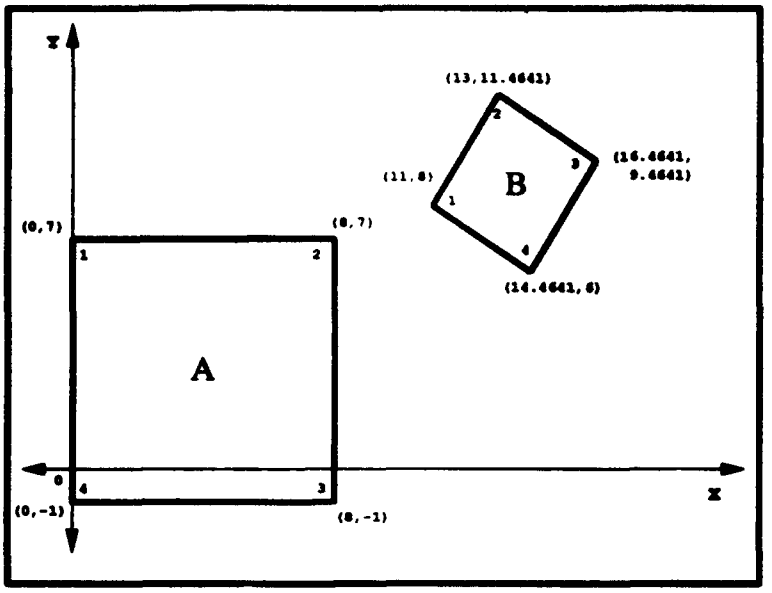


Figure 2.5 EDMA Example 2: Two objects of the same shape.

are similar to the first example. However, where object B has remained exactly the same, A has increased in size. The matrices of the (x, y) coordinates of the four landmarks of objects A and B are shown below:

$$\mathbf{A} = \begin{bmatrix} 0 & 7 \\ 8 & 7 \\ 8 & -1 \\ 0 & -1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 11 & 8 \\ 13 & 11.4641 \\ 16.4641 & 9.4641 \\ 14.4641 & 6 \end{bmatrix}$$

Again, the next step is creating the form matrix for each object where the distances between every pair of landmarks on each object are calculated. Because object B did not change, the calculation of $\mathbf{F(B)}$ is the same as in Example 1. However, the

FM of object A has changed. Both FMs are shown:

$$\mathbf{F(A)} = \begin{bmatrix} 0 & 8 & 11.3137 & 8 \\ 8 & 0 & 8 & 11.3137 \\ 11.3137 & 8 & 0 & 8 \\ 8 & 11.3137 & 8 & 0 \end{bmatrix}$$

$$\mathbf{F(B)} = \begin{bmatrix} 0 & 4 & 5.6569 & 4 \\ 4 & 0 & 4 & 5.6569 \\ 5.6569 & 4 & 0 & 4 \\ 4 & 5.6569 & 4 & 0 \end{bmatrix}$$

To compare A and B, the form difference matrix is calculated. Using **A** as the reference matrix, the $(i,j)^{\text{th}}$ element of the FDM is found by creating ratios of the $(i,j)^{\text{th}}$ elements of the form matrices above. For example, $D_{13}(\mathbf{A},\mathbf{B}) = \frac{F_{13}(\mathbf{A})}{F_{13}(\mathbf{B})} = \frac{11.3137}{5.6569} = 2$. The FDM of objects A and B is:

$$\mathbf{D(A,B)} = \begin{bmatrix} 0 & 2 & 2 & 2 \\ 2 & 0 & 2 & 2 \\ 2 & 2 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{bmatrix}$$

Note that the off-diagonal elements are equal to two. This indicates that the reference object A is the same shape as object B, but they differ in size: A is twice as large as B.

EDMA Example 3: Two different objects. In this example, object A is the same as in Example 2, but now, object B is altered; it is now four sided and non-rectangular. Again, the four corners of both objects represent landmarks,

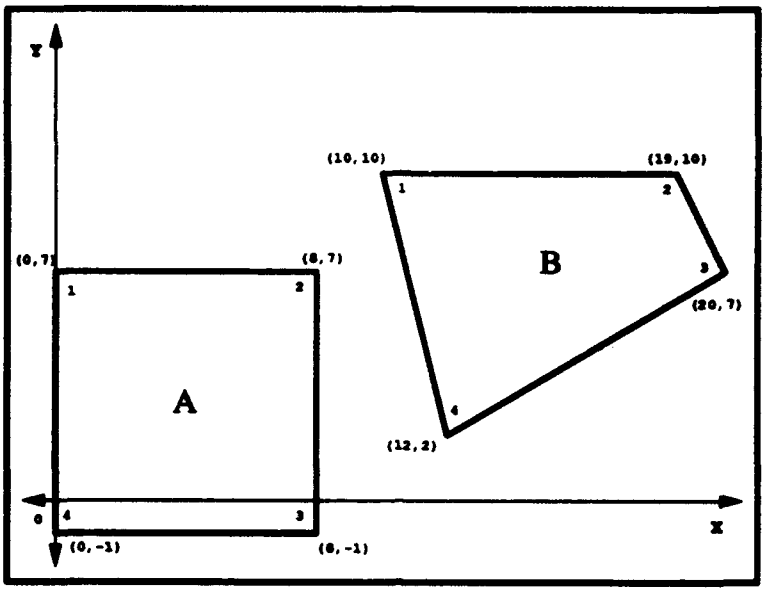


Figure 2.6 EDMA Example 3: Two different objects.

and the following matrices represent the (x, y) coordinates of the landmarks for objects A and B in Figure 2.6:

$$\mathbf{A} = \begin{bmatrix} 0 & 7 \\ 8 & 7 \\ 8 & -1 \\ 0 & -1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 10 & 10 \\ 19 & 10 \\ 20 & 7 \\ 12 & 2 \end{bmatrix}$$

Calculating the FMs for both objects is the next step. Because object A did not change, the calculation of $\mathbf{F(A)}$ is the same as in Example 2. However, the FM of object B has changed, and for convenience both FMs are shown:

$$\mathbf{F(A)} = \begin{bmatrix} 0 & 8 & 11.3137 & 8 \\ 8 & 0 & 8 & 11.3137 \\ 11.3137 & 8 & 0 & 8 \\ 8 & 11.3137 & 8 & 0 \end{bmatrix}$$

$$\mathbf{F(B)} = \begin{bmatrix} 0 & 9 & 10.4403 & 8.2462 \\ 9 & 0 & 3.1623 & 10.6301 \\ 10.4403 & 3.1623 & 0 & 9.4340 \\ 8.2462 & 10.6301 & 9.4340 & 0 \end{bmatrix}$$

To compare A and B, the form difference matrix is calculated. Using **A** as the reference matrix the FDM of objects A and B is:

$$\mathbf{D(A,B)} = \begin{bmatrix} 0 & 0.8889 & 1.0837 & 0.9701 \\ 0.8889 & 0 & 2.5298 & 1.0643 \\ 1.0837 & 2.5298 & 0 & 0.8480 \\ 0.9701 & 1.0643 & 0.8480 & 0 \end{bmatrix}$$

As the FDM shows, objects A and B are not the same form. The most significant change of distances between landmarks is between points 2 and 3. This is shown by the high value (2.5298) of element D_{23} (or D_{32} because of symmetry) in the FDM matrix.

2.2.3.4 Comparison of Multiple Objects. To compare more than two objects an overall reference object must be established. The reference object is compared to each target object one at a time. The objects are then related by how they differ from the reference object. Typically in morphometric studies the objects are pre-segregated and the reference object is calculated as the mean FM of the sample which contains the largest number of objects. An example of this type of pre-segregation is the study of skull growth in *Cebus apella* (a certain type of monkey) conducted by Corner and Richtsmeier (14).

In this study, the skulls were first segregated into six different age groups and then each age group was divided by gender (a total of 12 groups). To make comparisons within each age/gender group, the average or mean of the Euclidean

distances within each individual form matrix for each of the 12 groups was calculated. This resulted in 12 mean reference forms, one for each group. To compare all skulls to each other (no matter what age or gender), the reference object was chosen to be the mean form matrix from the group with the greatest number of objects.

2.2.3.5 EDMA Summary. Euclidean Distance Matrix Analysis is a type of morphometric analysis which is coordinate-free but landmark dependent. When comparing objects, the same number of landmarks must be present on each object, and there must be a one-to-one correspondence between the landmarks. EDMA can be used to compare multiple objects, but only relative to a reference object. When comparing objects within a sample, the reference form can be determined by pre-segregation of the sample and taking the mean of the group with the largest number of objects. The advantages of EDMA are the simple calculations involved and the coordinate-free property. Disadvantages include the dependency on landmarks and requirement of a reference object to compare multiple objects.

2.2.4 Thin Plate Spline. The thin plate spline (TPS) was first used for anthropometrics by Fred Bookstein (4, 5); the following section draws heavily from his work. The TPS is an algebraic method for describing the deformation of landmarks on a target object to corresponding landmarks on a reference object.

2.2.4.1 Bending Energy. A surface spline may be visualized by bending a flat thin steel plate to pass through each landmark of an object. The thin plate spline is the unique spline which requires the lowest physical bending energy for the plate to pass through all the points.

The bending energy is a function of the work done against the elasticity of the plate and not against gravity. Therefore, translation and rotation of the plate require no bending energy. The bending energy at any point is proportional to the

curvature of the plate at that point. At points not defined by landmarks, the plate is an approximation of the figure at that point.

If $z(x, y)$ is the height of the TPS, then the total curvature is found by integrating the second derivative of z over the entire plate. Minimizing the function:

$$\text{Total Curvature} = \iint_{\Re^2} \left[\left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 \right] dx dy \qquad (2.5)$$

describes the plate with the minimum required bending energy. The solution will give the final form of the plate. Recall that bending the plate requires energy; sharper bends have higher second derivatives and thus higher bending energies.

2.2.4.2 The $U(r)$ Function. The total curvature will be minimized if the biharmonic equation is zero. The biharmonic equation:

$$\begin{aligned} \Delta^2 U &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)^2 U \\ &= \left(\frac{\partial^2 U}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 U}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 U}{\partial y^2} \right)^2 \\ &= 0 \end{aligned}$$

describes the minimum bending energy steel plate. The plate is lofted in the z direction a distance $U = U(x, y)$ above a landmark on the x, y Cartesian plane. The fundamental solution is given by:

$$z(x, y) = -U(r) = -r^2 \ln r^2 \qquad (2.6)$$

where r is the Euclidean distance $\sqrt{x^2 + y^2}$, $z(x, y)$ is the height of the plate above the plane $z = 0$, and the negative sign is for convenience (4). Equation 2.6 is valid if and only if the object is made up of *one* landmark. When (x, y) is the landmark

coordinate, $z(x, y)$ is the height of the landmark. When (x, y) is not at the landmark, $z(x, y)$ is the estimated height of the object at that point.

The overall solution to the biharmonic equation is a linear combination of U at each landmark. In the overall solution, r_{i0} is the distance from the point being estimated (x_0, y_0) to the landmark at (x_i, y_i). For a figure described by n landmarks the overall solution is the spline:

$$z(x_0, y_0) = \hat{w}_1 U(r_{10}) + \hat{w}_2 U(r_{20}) + \cdots + \hat{w}_n U(r_{n0}) \tag{2.7}$$

where $\hat{w}_i$ are weighting factors or constants of integration. For convenience, the subscript '0' is often dropped.

2.2.4.3 Examples. The following examples illustrate the algebra of thin plate splines. For simplicity, the first two examples make assumptions about the form of the data. The general solution to the TPS problem is then found and illustrated in the third example.

2.2.4.4 TPS Example 1: Constant Displacements and Linear Regression to x, y Plane. This example illustrates the simplest possible thin plate spline: the landmarks are lofted some fixed distance above or below the x, y plane, and the ordinary least square regression of the data should then be the x, y plane.

One such data set is a square of side length $\sqrt{2}$ lofted a distance ± 2 at each vertex. The landmarks of the object are given by $(0, 1, -2)$, $(-1, 0, 2)$, $(0, -1, -2)$, and $(1, 0, 2)$.

Next, it is confirmed that the least square regression onto the arbitrary plane

$z = c_1 + c_2x + c_3y$ is indeed the x, y plane:

$$\begin{aligned}
\mathbf{AC} &= \mathbf{z} \\
\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} &= \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} \\
\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} &= \begin{bmatrix} -2 \\ 2 \\ -2 \\ 2 \end{bmatrix} \\
\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} &= (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{z} \\
\mathbf{C} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\end{aligned} \tag{2.8}$$

Therefore the linear regression gives $z = 0 + 0x + 0y = 0$ everywhere. The thin plate spline can now be calculated:

$$z(x, y) = \hat{w}_1 U(r_1) + \hat{w}_2 U(r_2) + \hat{w}_3 U(r_3) + \hat{w}_4 U(r_4) \tag{2.9}$$

where $\hat{w}_i$ are weighting factors to force the plate to pass through the data. In this example $\hat{w}_1 = -\hat{w}_2 = \hat{w}_3 = -\hat{w}_4 = \hat{w}$ since all landmarks are lofted by the same distance above and below the x, y plane. Substituting in the data gives:

$$\begin{aligned}
z(x, y) = &\left[U\left(\sqrt{x^2 + (y-1)^2}\right) - U\left(\sqrt{(x+1)^2 + y^2}\right)\right. \\
&\left. + U\left(\sqrt{x^2 + (y+1)^2}\right) - U\left(\sqrt{(x-1)^2 + y^2}\right)\right] \times \hat{w}
\end{aligned} \tag{2.10}$$

Evaluating the right hand side at the landmark $(1,0,2)$ and solving for $\hat{w}$ gives:

$$\begin{aligned} z(1,0) &= \left[U(\sqrt{2}) - U(\sqrt{4}) + U(\sqrt{2}) - U(0)\right] \times \hat{w} \\ 2 &= [2\ln 2 - 4\ln 4 + 2\ln 2 - 0] \times \hat{w} \\ \hat{w} &= -0.7213 \end{aligned}$$

where $\lim_{r^2 \to 0} r^2 \ln r^2 = 0$. With the weighting factors $\hat{\mathbf{w}}$ found above, the height of the plate at the following points (which are coincident with landmarks) is:

x	y	z
0	1	-2
-1	0	2
0	-1	-2
1	0	2

Therefore, the plate passes exactly through each landmark. Recall that the landmarks are points from a three-dimensional object. The TPS can also be used to estimate the shape of the object at points where landmark data is not known. For example, at the point $(0.5, 0.4)$ the estimated height of the figure is $z(0.5,0.4) = [U(\sqrt{0.61}) - U(\sqrt{2.41}) + U(\sqrt{2.21}) - U(\sqrt{0.41})]\hat{w} = (-0.3034) \times (-0.7213) = +0.2188$.

2.2.4.5 TPS Example 2: Variable Displacement. In the previous example the landmarks are lofted above the x, y plane by ± 2. In this example the restriction of fixed lofting distance is removed. The first step is to calculate a spline which passes exactly through all landmarks. The second step shows that the spline is the minimum bending energy spline (i.e., the TPS). The data considered is $(0,1,1)$, $(-1,0,2.4188)$, $(0,-1,1)$, and $(3,0,2.5463)$.

The TPS is given by:

$$\begin{aligned} z(x,y) &= \hat{w}_1\, U\left(\sqrt{x^2+(y-1)^2}\right) + \hat{w}_2\, U\left(\sqrt{(x+1)^2+y^2}\right) \\ &\quad +\hat{w}_3\, U\left(\sqrt{x^2+(y+1)^2}\right) + \hat{w}_4\, U\left(\sqrt{(x-3)^2+y^2}\right) \end{aligned} \tag{2.11}$$

The four weighting factors can be found from the fact that the plate must pass exactly through the four landmarks. Solving the four simultaneous equations for the weighting factors gives:

$$\begin{bmatrix} z(x_1,y_1) \\ z(x_2,y_2) \\ z(x_3,y_3) \\ z(x_4,y_4) \end{bmatrix} = \begin{bmatrix} U(r_{11}) & U(r_{12}) & U(r_{13}) & U(r_{14}) \\ U(r_{21}) & U(r_{22}) & U(r_{23}) & U(r_{24}) \\ U(r_{31}) & U(r_{32}) & U(r_{33}) & U(r_{34}) \\ U(r_{41}) & U(r_{42}) & U(r_{43}) & U(r_{44}) \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \end{bmatrix}$$

$$\begin{bmatrix} z(0,1) \\ z(-1,0) \\ z(0,-1) \\ z(3,0) \end{bmatrix} = \begin{bmatrix} 0 & U(\sqrt{2}) & U(\sqrt{4}) & U(\sqrt{10}) \\ U(\sqrt{2}) & 0 & U(\sqrt{2}) & U(\sqrt{16}) \\ U(\sqrt{4}) & U(\sqrt{2}) & 0 & U(\sqrt{10}) \\ U(\sqrt{10}) & U(\sqrt{16}) & U(\sqrt{10}) & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 2.4118 \\ 1 \\ 2.5463 \end{bmatrix} = \begin{bmatrix} 0 & 2\ln 2 & 4\ln 4 & 10\ln 10 \\ 2\ln 2 & 0 & 2\ln 2 & 16\ln 16 \\ 4\ln 4 & 2\ln 2 & 0 & 10\ln 10 \\ 10\ln 10 & 16\ln 16 & 10\ln 10 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \end{bmatrix}$$

$$\begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \end{bmatrix} = \begin{bmatrix} -0.1243 \\ 0.1864 \\ -0.1243 \\ 0.0621 \end{bmatrix} \tag{2.12}$$

The next step is to show that the spline (Eq. 2.11) along with the constants from Equation 2.12 relate to the minimum bending energy spline. Equation 2.5 gives the total curvature (i.e., bending energy) of the plate. From Bookstein (5:pp. 33–

34) the total curvature is the sum of the curvature in the x and y directions taken independently and is minimized if $\sum \hat{w} = \sum \hat{w}x = \sum \hat{w}y = 0$. For the above data:

$$\sum \hat{w} = -0.1243 + 0.1864 - 0.1243 + 0.0621 = 0 \quad (2.13)$$

$$\sum \hat{w}x = (-0.1243)(0) + (0.1864)(-1) - (0.1243)(0) + (0.0621)(3) = 0 \quad (2.14)$$

$$\sum \hat{w}y = (-0.1243)(1) + (0.1864)(0) - (0.1243)(-1) + (0.0621)(0) = 0 \quad (2.15)$$

showing that the spline is the minimum bending energy TPS.

2.2.4.6 General Thin Plate Spline Equations. In this section the general equations for a thin plate spline are developed. In the general case, the least squares regression of the data is not necessarily the x, y plane. As stated earlier, the bending energy of a thin plate spline should be independent of translation and rotation of the entire plate. To account for this, the general TPS equations determine an optimal plane through the data and then calculate the deviation from that plane.

In the previous examples, the thin plate spline is given by:

$$z(x, y) = \sum_{\text{landmarks}} U(r_i)\hat{w}_i \quad (2.16)$$

which is the height the plate is lofted above the x, y plane. For a general data set, the data might not deviate only from the z-plane. The data could contain an average tendency described by the plane:

$$z = c_1 + c_2 x + c_3 y \quad (2.17)$$

The TPS is then:

$$z(x, y) = c_1 + c_2 x + c_3 y + \sum_{\text{landmarks}} U(r_i)\hat{w}_i \quad (2.18)$$

The thin plate spline is a valid interpolate of the data for any point (x, y). However, there are $n+3$ unknowns $(c_1, c_2, c_3, \hat{w}_1, \ldots, \hat{w}_n)$ which must be calculated. The first n equations come from $z = z(x, y)$ which is known at the n landmarks. The last three equations come from $\sum \hat{w} = \sum \hat{w}x = \sum \hat{w}y = 0$. The equations are:

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \nwarrow & \uparrow & \nearrow & 1 & x_1 & y_1 \\ \leftarrow & \mathbf{K} & \rightarrow & \vdots & \vdots & \vdots \\ \swarrow & \downarrow & \searrow & 1 & x_n & y_n \\ 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \vdots \\ \hat{w}_n \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{2.19}$$

where:

$$\mathbf{K} = \begin{bmatrix} 0 & U(r_{12}) & U(r_{13}) & \cdots & U(r_{1n}) \\ U(r_{21}) & 0 & U(r_{23}) & \cdots & U(r_{2n}) \\ U(r_{31}) & U(r_{32}) & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & U(r_{n-1,n}) \\ U_{n1}) & U(r_{n2}) & \cdots & U(r_{n,n-1}) & 0 \end{bmatrix} \tag{2.20}$$

These equations can be written in a more compact form by defining:

$$\mathbf{P} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \tag{2.21}$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \tag{2.22}$$

$$\mathbf{y} = [\hat{w}_1 \quad \cdots \quad \hat{w}_n \quad c_1 \quad c_2 \quad c_3]^{\mathrm{T}} \tag{2.23}$$

$$\hat{\mathbf{z}} = [z_1 \quad \cdots \quad z_n \quad 0 \quad 0 \quad 0]^{\mathrm{T}} \tag{2.24}$$

then:

$$\hat{\mathbf{z}} = \mathbf{L}\mathbf{y} \tag{2.25}$$

where $\mathbf{0}$ is a 3×3 matrix of zeros. Since all the elements of $\mathbf{L}$ and $\hat{\mathbf{z}}$ are known for a given data set, the constants $\mathbf{y}$ are calculated from:

$$\mathbf{y} = \mathbf{L}^{-1}\hat{\mathbf{z}} \tag{2.26}$$

With the constants calculated, the position of the figure at any point can be interpolated from the thin plate spline (Eq. 2.18).

It is important to note that $\mathbf{L}_n^{-1}$, the upper left $n \times n$ elements of $\mathbf{L}^{-1}$, is the bending energy matrix. It represents the energy required to deform a flat plate to the data set under the assumption that $U(r) = r^2 \ln r^2$ (5:p. 319). For the general thin plate spline, the bending energy matrix $\mathbf{L}_n^{-1}$ completely describes the form of the figure.

2.2.4.7 TPS Example 3: General TPS Calculation. In this example the thin plate spline is found for a general set of data. The data set considered is $(2,4,7)$, $(3,10,2)$, $(-4,7,1)$, $(7,-2,-4)$, and $(4,1,5)$. The spline constants are calculated from Equation 2.25:

$$\begin{bmatrix} z(2,4) \\ z(3,10) \\ z(-4,7) \\ z(7,-2) \\ z(4,1) \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & U(\sqrt{37}) & U(\sqrt{45}) & U(\sqrt{61}) & U(\sqrt{13}) & 1 & 2 & 4 \\ U(\sqrt{37}) & 0 & U(\sqrt{58}) & U(\sqrt{160}) & U(\sqrt{82}) & 1 & 3 & 10 \\ U(\sqrt{45}) & U(\sqrt{58}) & 0 & U(\sqrt{202}) & U(\sqrt{100}) & 1 & -4 & 7 \\ U(\sqrt{61}) & U(\sqrt{160}) & U(\sqrt{202}) & 0 & U(\sqrt{18}) & 1 & 7 & -2 \\ U(\sqrt{13}) & U(\sqrt{82}) & U(\sqrt{100}) & U(\sqrt{18}) & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & -4 & 7 & 4 & 0 & 0 & 0 \\ 4 & 10 & 7 & -2 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \\ \hat{w}_5 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

$$\begin{bmatrix} 7 \\ 2 \\ 1 \\ -4 \\ 5 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 37\ln 37 & 45\ln 45 & 61\ln 61 & 13\ln 13 & 1 & 2 & 4 \\ 37\ln 37 & 0 & 58\ln 58 & 160\ln 160 & 82\ln 82 & 1 & 3 & 10 \\ 45\ln 45 & 58\ln 58 & 0 & 202\ln 202 & 100\ln 100 & 1 & -4 & 7 \\ 61\ln 61 & 160\ln 160 & 202\ln 202 & 0 & 18\ln 18 & 1 & 7 & -2 \\ 13\ln 13 & 82\ln 82 & 100\ln 100 & 18\ln 18 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & -4 & 7 & 4 & 0 & 0 & 0 \\ 4 & 10 & 7 & -2 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \\ \hat{w}_5 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

$$\begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \\ \hat{w}_4 \\ \hat{w}_5 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0.0326 \\ -0.0101 \\ -0.0240 \\ -0.0457 \\ 0.0471 \\ 22.121 \\ -0.3381 \\ 0.2234 \end{bmatrix} \tag{2.27}$$

The height of the plate at any (x, y) coordinate can now be calculated from the thin plate spline (Eq. 2.18):

$$\begin{aligned} z(x,y) = \; & \hat{w}_1 U\left(\sqrt{(x-2)^2+(y-4)^2}\right) + \hat{w}_2 U\left(\sqrt{(x-3)^2+(y-10)^2}\right) \\ & + \hat{w}_3 U\left(\sqrt{(x+4)^2+(y-7)^2}\right) + \hat{w}_4 U\left(\sqrt{(x-7)^2+(y+2)^2}\right) \\ & + \hat{w}_5 U\left(\sqrt{(x-4)^2+(y-1)^2}\right) + c_1 + c_2 x + c_3 y \end{aligned} \tag{2.28}$$

2.2.4.8 Comparison of Objects. Up to this point, thin plate splines have been used to interpolate the coordinates of one object. This interpolation can be thought of as shifting the z coordinate of the object from $z = 0$ to a nonzero value.

An equivalent way to think of the z shift is to consider a reference object with coordinates $(x_i, y_i, 0)$ and a target object with coordinates (x_i, y_i, z_i). In this case the x shift and the y shift are zero since the x and y coordinates of the reference and target objects are the same. The z shift is a measure of the energy required to bend a thin plate through the reference object to a thin plate through the target object.

When interpolating an object, the same x and y coordinates are used for the 'reference' and 'target' object and the z shift is calculated. The z shift represents the difference between the reference and target objects. Similarly, differences in x and y coordinates can be used to calculate the x shift and the y shift.

For two-dimensional objects, the x shift and the y shift completely describe the differences between the reference and target objects. For three-dimensional objects, the x shift, the y shift, and the z shift completely describe the differences between the reference and target objects. The following examples demonstrate how the shifts are calculated and interpreted.

2.2.4.9 TPS Example 4: Objects of Identical Form. This example calculates the x shift and y shift for two two-dimensional objects. The reference object has coordinates $(2,4)$, $(3,10)$, $(-4,7)$, $(7,-2)$, and $(4,1)$ which are the (x,y) coordinates considered in TPS Example 3. The target coordinates are $(3.83013, -3.36603)$, $(7.69615, 1.33013)$, $(0.133975, 2.23205)$, $(5.16025, -11.0622)$, and $(4.06218, -6.9641)$.

The target object is the reference object translated $+3$ units in the x direction, translated -5 units in the y direction, and rotated $+30$ degrees about the z - axis. Therefore, the reference and target object have the same form.

By analogy to Equation 2.19, the constants for the x shift are given by:

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_n \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \nwarrow & \uparrow & \nearrow & 1 & x_1 & y_1 \\ \leftarrow & \mathbf{K} & \rightarrow & \vdots & \vdots & \vdots \\ \swarrow & \downarrow & \searrow & 1 & x_n & y_n \\ 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{x1} \\ \vdots \\ \hat{w}_{xn} \\ c_{x1} \\ c_{x2} \\ c_{x3} \end{bmatrix} \tag{2.29}$$

where $\mathbf{x}'$ are the x coordinates of the target object, $\mathbf{L}$ is calculated from the reference object, and $\hat{\mathbf{w}}_x$ and $\mathbf{c}_x$ are the weights for the x shift. Substituting in the coordinates

of the objects gives:

$$\begin{bmatrix} 3.83013 \\ 7.69615 \\ 0.133975 \\ 5.16025 \\ 4.06218 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 37\ln 37 & 45\ln 45 & 61\ln 61 & 13\ln 13 & 1 & 2 & 4 \\ 37\ln 37 & 0 & 58\ln 58 & 160\ln 160 & 82\ln 82 & 1 & 3 & 10 \\ 45\ln 45 & 58\ln 58 & 0 & 202\ln 202 & 100\ln 100 & 1 & -4 & 7 \\ 61\ln 61 & 160\ln 160 & 202\ln 202 & 0 & 18\ln 18 & 1 & 7 & -2 \\ 13\ln 13 & 82\ln 82 & 100\ln 100 & 18\ln 18 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & -4 & 7 & 4 & 0 & 0 & 0 \\ 4 & 10 & 7 & -2 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{x1} \\ \hat{w}_{x2} \\ \hat{w}_{x3} \\ \hat{w}_{x4} \\ \hat{w}_{x5} \\ c_{x1} \\ c_{x2} \\ c_{x3} \end{bmatrix}$$

$$\begin{bmatrix} \hat{w}_{x1} \\ \hat{w}_{x2} \\ \hat{w}_{x3} \\ \hat{w}_{x4} \\ \hat{w}_{x5} \\ c_{x1} \\ c_{x2} \\ c_{x3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.098076 \\ 0.866025 \\ 0.5 \end{bmatrix} \tag{2.30}$$

Substituting the constants into the TPS equation gives the x shift:

$$f_x(x, y) = 0.098076 + 0.866025x + 0.5y \tag{2.31}$$

where x and y are the coordinates of the reference object. Similarly, the y shift is:

$$f_y(x, y) = -5.83013 - 0.5x + 0.866025y \tag{2.32}$$

The x and y shifts are flat plates with no bending. Since rotating and translating a plate requires no energy, there is no energy required to shift the reference object to the target object. Therefore the two objects have the same form.

The shifts can be used to estimate the coordinates of the target object. At the reference object landmark $(2, 4)$, the estimated x and y shifts are $f_x(2, 4) = 3.83013$ and $f_y(2, 4) = -3.36603$ which is coincident with the first landmark of the target object. Similarly, all the landmarks of the reference object shift to the equivalent landmark of the target object.

The shifts can also be used to interpolate points between landmarks. A point at $(5, 8)$ on the reference object is equivalent to $(f_x, f_y) = (8.4282, -1.40192)$ on the target object. The next example compares objects of different form.

2.2.4.10 TPS Example 5: Objects of Different Form. This example explains the comparison of objects with different form. The reference object is the same two-dimensional object as the previous example. The target object is represented by the points $(6, -5)$, $(3, -1)$, $(-3, 4)$, $(9, -10)$, and $(7, -4)$. The x shift constants are given by:

$$\begin{bmatrix} 6 \\ 3 \\ -3 \\ 9 \\ 7 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 37\ln 37 & 45\ln 45 & 61\ln 61 & 13\ln 13 & 1 & 2 & 4 \\ 37\ln 37 & 0 & 58\ln 58 & 160\ln 160 & 82\ln 82 & 1 & 3 & 10 \\ 45\ln 45 & 58\ln 58 & 0 & 202\ln 202 & 100\ln 100 & 1 & -4 & 7 \\ 61\ln 61 & 160\ln 160 & 202\ln 202 & 0 & 18\ln 18 & 1 & 7 & -2 \\ 13\ln 13 & 82\ln 82 & 100\ln 100 & 18\ln 18 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & -4 & 7 & 4 & 0 & 0 & 0 \\ 4 & 10 & 7 & -2 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{x1} \\ \hat{w}_{x2} \\ \hat{w}_{x3} \\ \hat{w}_{x4} \\ \hat{w}_{x5} \\ c_{x1} \\ c_{x2} \\ c_{x3} \end{bmatrix}$$

$$\begin{bmatrix} \hat{w}_{x1} \\ \hat{w}_{x2} \\ \hat{w}_{x3} \\ \hat{w}_{x4} \\ \hat{w}_{x5} \\ c_{x1} \\ c_{x2} \\ c_{x3} \end{bmatrix} = \begin{bmatrix} 0.033978 \\ -0.006510 \\ -0.009051 \\ -0.003654 \\ -0.014762 \\ 9.23548 \\ 0.912579 \\ -0.307918 \end{bmatrix} \tag{2.33}$$

The x shift is given by:

$$\begin{aligned} f_x(x_0, y_0) &= 0.033978\, U(r_{10}) - 0.006510\, U(r_{20}) - 0.009051\, U(r_{30}) - 0.003654\, U(r_{40}) \\ &\quad -0.014762\, U(r_{50}) + 9.23548 + 0.912579 x_0 - 0.307918 y_0 \end{aligned} \tag{2.34}$$

where (x_0, y_0) is a point on the reference object and f_x is the x coordinate of the equivalent point on the target object. The y shift is given by:

$$\begin{aligned} f_y(x_0, y_0) &= -0.079285\ U(r_{10}) + 0.011934\ U(r_{20}) + 0.008095\ U(r_{30}) - 0.027290\ U(r_{40}) \\ &\quad +0.086546\ U(r_{50}) - 4.33914 - 1.06199 x_0 + 0.609857 y_0 \end{aligned} \tag{2.35}$$

At the first landmark of the reference object:

$$\begin{aligned} f_x(2,4) &= 0 - 0.006510(37 \ln 37) - 0.009051(45 \ln 45) - 0.003654(61 \ln 61) \\ &\quad -0.014762(13 \ln 13) + 9.23548 + (0.912579)(2) - (0.307918)(4) \\ &= 6 \\ f_y(2,4) &= 0 + 0.011935(37 \ln 37) + 0.008095(45 \ln 45) - 0.0272904(61 \ln 61) \\ &\quad +0.086546(13 \ln 13) - 4.33914 + (-1.06199)(2) + (0.609857)(4) \\ &= -5 \end{aligned}$$

Therefore, the thin plate does convert the landmarks of the reference object to the landmarks of the target object.

The affine portion of the shifts orients the object in procrustes optimal position. The non-affine part of the x and y shifts are:

$$\begin{aligned} f_{\mathrm{na}_x}(x_0, y_0) &= 0.033978\ U(r_{10}) - 0.006510\ U(r_{20}) - 0.009051\ U(r_{30}) \\ &\quad -0.003654\ U(r_{40}) - 0.014762\ U(r_{50}) \end{aligned} \tag{2.36}$$

$$\begin{aligned} f_{\mathrm{na}_y}(x_0, y_0) &= -0.079285\ U(r_{10}) + 0.011934\ U(r_{20}) + 0.008095\ U(r_{30}) \\ &\quad -0.027290\ U(r_{40}) + 0.086546\ U(r_{50}) \end{aligned} \tag{2.37}$$

These represent the true differences in form as a function of the (x, y) location on the reference object. They are a measure of the energy to bend a plate passing through the reference object to a plate passing through the target object.

The shifts can be calculated between the reference object and any target object with the equivalent five landmarks. By comparing these shifts, differences in the form of objects can be compared.

2.2.4.11 TPS Summary. This section used several examples to develop the algebra of thin plate splines. The spline can be thought of as a plate bent to pass through all the landmarks of a three-dimensional object. While there are an infinite number of splines that will pass through all the landmarks, the thin plate spline is the spline which requires the minimum bending or curvature of the plate.

Thin plate splines can also be used to compare the difference between objects. This comparison has many advantages to regression based techniques; however, since landmarks need to be specified, this method is inappropriate for comparing the facial seal lines of oxygen masks.

In the examples, the bending energy matrix to deform a flat plate is exploited to find the form of the object and to compare the form of two objects. There are several advantages to using thin plate splines to compare objects. Objects are compared by bending a plate from one object to another object. Thin plate splines give both the magnitude and direction of stretch between the objects while other methods provide only scalar difference between objects. Thin plate splines put the objects into procrustes-optimal position saving a step in the calculations. Finally (and most importantly), Thin plate splines account for the distance between landmarks while regression based methods ignore the relationship between a landmark and its neighbors (5).

There are two major disadvantages to comparing objects using the thin plate spline method. The first is that thin plate splines is landmark dependent. Landmarks must be predetermined and correlated from object to object. The second is the need to define a 'standard' object from which all other objects are compared.

2.2.5 Form Analyses Summary: Procrustes, FESA, EDMA, TPS. The form analytical methods of this section identify methods currently available for form description and comparison. Many of these have been specifically developed for application to anthropometrics. However, each method is unique. Procrustes defines a scalar which combines the translation, rotation, size, and shape differences between faces. Finite element scaling analysis measures the 'strain' required to bend one face to another. Euclidean distance matrix method measures the relative distance between landmarks of different faces. Thin plate splines measure the energy to bend a flat plate from a reference object to a target object. A drawback to these four methods is the need to define a reference object. Pre-segregation of objects, such as the method discussed in Section 2.2.3, to determine a reference object is not feasible in this research. It is one goal of this research to cluster the faces into groups – not to pre-segregate them. Another major disadvantage is the use of landmarks. All four of these methods require the use of landmarks.

Landmarks are common points for obtaining specific relationships or identifying growth patterns. However, landmarks can be hard to locate without the experience and expertise of someone familiar with anthropometrics. Landmarks like the *sellion* (indentation in the bridge of the nose at the brow) and *promenton* (high point on the chin) may not be uniquely defined on certain faces due to ethnic backgrounds or other dissimilarities (see Section 5.17). Typically, a reference plane is established, like the Frankfort Horizontal discussed in Section 5.1.2.2, so common features are easily referenced (34). However, a reference of this type requires the entire head of an individual as the source of data. The data obtained for this research consist of plaster casts of the face only. Landmarks also require one-to-one correspondence between objects, so the data sets must be equal in size. It is important to note that landmarks alone do not provide all information pertaining to the form of an object. When using landmarks, information on curvature and surface features between landmarks is lost (46). Therefore, landmarks neglect information regarding

the relative shape between landmarks, which may be crucial in the determination of the mask-to-face fit. In essence, the geometry of the human face is not easily described by landmarks. The line of contact between the oxygen mask seal and the wearer's face is a complex, three-dimensional curve. It varies widely from face to face, and the selection of landmarks that would adequately capture the geometry of every face would be virtually impossible.

2.3 Conclusions

Past Air Force morphometric analyses, such as the multivariate approaches discussed in Section 2.1, could not reconstruct biological objects from the linear dimensions collected. Linear dimensions were measured between landmarks thought to distinguish different characteristics of the biological object. Past analyses usually resulted in a series of traits that described particular features but did not define the geometry of the object (46). Therefore, it is necessary to determine a form characterization method which captures all the critical features necessary for comparison and can reconstruct the form of the object under study.

Several morphological methods are available for form characterization and comparison. However, procrustes, finite element scaling analysis, Euclidean distance matrix analysis, and thin plate splines cannot be used to compare the facial seal lines of aircrew oxygen masks because they are landmark dependent and do not provide an accurate surface description.

III. Form Characterization

3.1 Introduction

The form characterization methods described in Chapter II do not capture the true shape or curvature of the face for description, but compare linear distances requiring the use of facial landmarks. It is the intent of this research to develop a form characterization method which can accurately describe the curvature of the face in a form compatible with clustering. This chapter describes a form characterization methodology which determines the form of a face by utilizing the spacial correlation of the surface data. It does not require the use of landmarks and is applicable to unequal size data sets. This new methodology results in an energy matrix, the energy to transform a flat plane into a face. This matrix is unique to each individual face, and is the source for feature selection in clustering.

The energy matrix employs many concepts from kriging, a geostatistical method typically used for predicting ore content at unsampled locations. Therefore, a brief discussion on kriging precedes the explanation on the development of the energy matrix.

3.2 Kriging

The term 'kriging' was introduced by G. Matheron (1963) to describe a spacial prediction method, named after D.G. Krige, a South African mining engineer, who in 1951 had developed empirical methods for determining ore reserves based on distributions of sampled data (17:p. 199). D.G. Krige was a pioneer in applying statistical techniques to mine evaluation, but it was through Matheron's work that the field of geostatistics was born. Geostatistics is the application of statistics to geology, of which kriging is a part.

Kriging can be described as the probabilistic process of obtaining the best linear unbiased predictor (BLUP) for an unknown variable, where 'best' is defined

as having minimum estimation error variance (18). It was originally used to predict ore reserves based on a few test samples. Kriging is unique in that it combines spacial correlation with linear unbiased estimators. The outcome of kriging is a system of equations that provides the optimum estimation of an unknown condition, be it elevation, ore content, or some other desired variable at an unsampled location. The kriging system is also known as an exact interpolator because it honors the data (20); i.e., it returns the measured values at all sampled locations.

Before kriging can be applied, a thorough structure analysis must be completed. Structure analysis is the development of the spacial structure or correlation in the data through the identification of the variogram. The structure analysis includes estimation of the trend or drift in the data, development of a functional variogram model, and identification of anisotropy. After the structure analysis is completed then the kriging system can be applied and the kriging variance, or minimum estimation variance can be calculated.

3.2.1 Variogram. A variogram describes the spacial variability of a data set. It is a function solely dependent on $\mathbf{h}$, the distance and orientation between two locations. As stated previously, the first step towards developing a kriging system for a multidimensional data set is the estimation of the variogram, $\gamma(\mathbf{h})$. The experimental variogram is first calculated from the data and then a functional variogram is found which best describes the spacial correlation within the data.

3.2.1.1 Assumptions. There are three major assumptions made when using the variogram. The first is that the variable to be modeled is regionalized. Unlike a random field, a regionalized variable can be used to describe a random function which has continuity from point to point, but changes in ways that are so complex they cannot be described by a simple combination of a deterministic function and a random variable. The second assumption is that the data is isotropic, and the third requires that the regionalized variable exhibit at least intrinsic stationarity.

Stationarity refers to the moments of the regionalized variable (19), and is a model decision, not an inherent property of the regionalized variable (31).

There are three types of stationarity: weak, strong, and intrinsic. Weak stationarity, also called wide-sense or second-order stationarity, relates only to the first two moments of the regionalized variable $z(\mathbf{x})$ where $\mathbf{x}$ is the location (x, y), and $z(\mathbf{x})$ is the variable of interest at $\mathbf{x}$ (e.g., elevation or ore content). The definition of weak stationarity requires that the expected value of the regionalized variable, $\mathrm{E}[z(\mathbf{x})]$, be finite and the same for all $\mathbf{x}$, and the spacial covariance of the regionalized variable, $\mathrm{C}(\mathbf{h})$, is also finite and the same for all $\mathbf{x}$. If the expected value is $\mathrm{E}[z(\mathbf{x})] = \mu$, then the covariance is:

$$\begin{aligned} \mathrm{E}[(z(\mathbf{x}) - \mu)(z(\mathbf{x}+\mathbf{h}) - \mu)] &\equiv \mathrm{C}[z(\mathbf{x}+\mathbf{h}), z(\mathbf{x})] \\ &= \mathrm{C}(\mathbf{h}) \end{aligned} \tag{3.1}$$

where $z(\mathbf{x})$ is a measurement of the regionalized variable taken at a specific location $\mathbf{x}$, and $z(\mathbf{x}+\mathbf{h})$ is another measurement taken $\mathbf{h}$ distance away. Note that $\mathbf{h}$ is a vector, and h is the magnitude of $\mathbf{h}$, the two-dimensional Euclidean distance between two locations:

$$h_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3.2}$$

where $i = 1, ..., n$, $j = 1, ..., n$, and n is the total number of locations in the data set. The variance of the regionalized variable $z(\mathbf{x})$ is:

$$\begin{aligned} \mathrm{Var}[z(\mathbf{x})] &= \mathrm{E}[(z(\mathbf{x}) - \mu)^2] \\ &= \mathrm{C}(0) \end{aligned} \tag{3.3}$$

Therefore, the covariance only exists if $\mathrm{Var}[z(\mathbf{x})]$ is finite.

Strong stationarity, also known as strict-sense stationarity, includes higher order moments of the regionalized variable, and exists when all the moments are de-

fined and finite. Strong stationarity implies the existence of weak stationarity, but the reverse is not true.

Intrinsic stationarity also relates to the first two moments but applies to situations where the variance of the regionalized variable is not finite. The underlying assumption is that the variance of the changes in the regionalized variable over the data set is finite, and can be defined. The first two moments of the variations in the regionalized variable under intrinsic stationarity conditions can be defined as:

$$\mathrm{E}[z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})] = 0 \tag{3.4}$$

$$\mathrm{Var}[z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})] = 2\gamma(\mathbf{h}) \tag{3.5}$$

where $\gamma(\mathbf{h})$ is defined to be the variogram. Notice that intrinsic stationarity is a generalization of weak stationarity, where the mean is zero over the data set. Given weak stationarity conditions, the relationship between the variogram to the covariance is:

$$\gamma(\mathbf{h}) = \mathrm{C}(0) - \mathrm{C}(\mathbf{h}) \tag{3.6}$$

If the regionalized variable is normalized to have a maximum variance of 1.0, the variogram $\gamma(\mathbf{h})$ is a mirror image of the corresponding autocorrelation function, Figure 3.1 (20). Given these relationships, a variogram representing the spacial correlation of the regionalized variable is estimated in order to determine the coefficients for an optimal linear predictor in kriging (19:pp. 92–93).

3.2.1.2 Definition. The variogram is essentially the variance of the data as a function of $\mathbf{h}$, the distance or lag between data points. Recalling the intrinsic stationarity assumption, the variogram can be defined for a zero mean

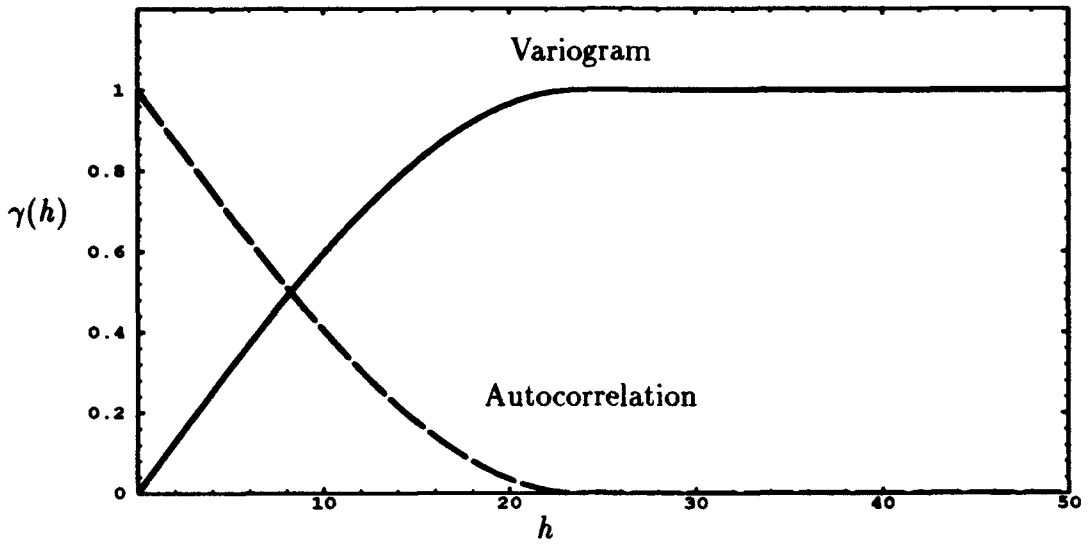


Figure 3.1 Relationship between the variogram and autocorrelation versus separation distance, h.

random function:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= \mathrm{Var}[z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})] \\ &= \mathrm{E}\left\{[z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})]^2\right\} \end{aligned} \tag{3.7}$$

For example, choosing the elevation or height as a function of location, $z(\mathbf{x})$, the variance in height can be estimated using the mean squared differences between data points, a given distance h apart in a specific direction:

$$2\gamma(\mathbf{h}) = \frac{1}{N(\mathbf{h})}\sum_{i=1}^{n}[z(\mathbf{x}_i+\mathbf{h}) - z(\mathbf{x}_i)]^2 \tag{3.8}$$

where $N(\mathbf{h})$ is the number of samples separated by a specific $\mathbf{h}$. This relationship is based on the assumption that the difference in value between two positions depends only on the distance between them and their relative orientation $\mathbf{h}$. The '2' preceding the $\gamma(\mathbf{h})$ is there for mathematical convenience and is a consequence of the fact that the focus is on the perpendicular distance of the points from the 45-degree line. Isaaks, *et al*, (29) equate the variogram to the moment of inertia, noting that as h

increases, the points drift away from the line where $\mathbf{h} = (0, 0)$; therefore, the moment of inertia about the 45-degree line is a natural measure. The result is:

$$\gamma(\mathbf{h}) = \gamma(-\mathbf{h}) \tag{3.9}$$

which indicates that the variogram calculated in one direction is equal to the variogram calculated in the opposite direction. Technically, the term $2\gamma(\mathbf{h})$ is called the variogram, and $\gamma(\mathbf{h})$ the semi-variogram, although some authors call the latter the variogram. In this discussion, $\gamma(\mathbf{h})$ is referred to as the variogram.

3.2.1.3 Functional Variogram. The variogram is considered to be the fundamental tool in geostatistics. It provides answers to basic questions regarding: the continuity of data, the direction of continuity, the area of influence (range), and the data isotropy/anisotropy (33:p. 84). As h increases and the data points appear farther apart, the correlation between the points decreases. The expectation is that when the distance becomes relatively large then the sample values become independent of one another. The variogram value becomes essentially constant, since it represents the relationship between sets of independent samples. The distance h at which samples become independent of one another is denoted by a and referred to as the 'range of influence' of a sample. This range or span, a, defines a neighborhood within which data have statistically significant correlation. In addition, the slope of the variogram in the neighborhood of $h = 0$ is an indication of the continuity of the data.

3.2.1.4 Variogram Models. The next step is determining a continuous model which provides the best fit to the variogram estimates. A preferred model is one which incorporates the three parameters of the ideal variogram: the range of influence, a; the sill, $(Co + C)$; and the nugget effect, Co (see Figure 3.2). The value of $\gamma(\mathbf{h})$ at which the graph levels off (at $h = a$) is denoted by $(Co + C)$ and is called

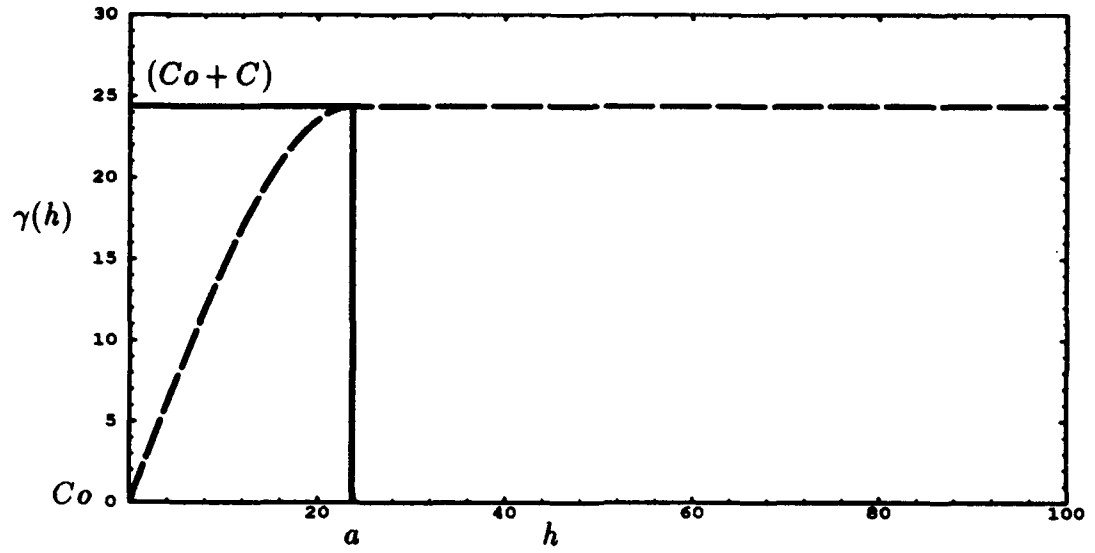


Figure 3.2 Variogram Parameters: range of influence, a; sill, $Co + C$; and nugget effect, Co.

the sill of the variogram (10:p. 6). The sill is equal to the sample variance of the data. The nugget effect accounts for sampling or measurement errors and is also an indication of the small-scale variability in the data (33). There are various methods available for deriving a, C, and Co. The simplest is using weighted least squares.

The selection of the underlying model is a crucial one. The model determines the continuous relationship of the points within the data, which in turn can affect the success of the kriging system to predict the unknown variable. However, the experimental variogram should not be overfit to the data. A simple model may provide solutions that are as accurate as those found using a more complex model. Therefore, the goal is to define a parsimonious model which includes these critical parameters. There are three models which are most common: spherical, exponential, and Gaussian (Fig 3.3).

Spherical Model. The spherical model, also known as the Matheron model, is the most commonly used model. It has linear behavior at small separation distances, but flattens out at larger h, reaching the sill at a. The tangent

at the origin reaches the sill at approximately $\frac{2a}{3}$:

$$\gamma(h) = \begin{cases} 0 & \text{when } h = 0 \\ Co + C\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) & \text{when } 0 < h < a \\ Co + C & \text{when } h \geq a \end{cases} \tag{3.10}$$

Exponential Model. The exponential model is asymptotic, with the variogram never reaching the sill. The exponential model, like the spherical model, is linear at very short distances near the origin; however, it rises more steeply, then flattens out more gradually. The tangent at the origin reaches the sill at approximately $\frac{a}{5}$:

$$\gamma(h) = Co + C\left[1 - \exp\left(\frac{-h}{a}\right)\right] \tag{3.11}$$

Gaussian Model. The Gaussian model reaches its sill asymptotically, and the parameter a is approximately the practical range or distance at which the variogram value is 95% of the sill. The distinguishing feature of the Gaussian model is its parabolic behavior near the origin; it is the only transition model presented whose shape has an inflection point:

$$\gamma(h) = Co + C\left[1 - \exp\left(\frac{-3h^2}{a^2}\right)\right] \tag{3.12}$$

The functional variogram can be used to indicate the presence of anisotropy in the data. If anisotropy exists, it must be corrected so the data can be considered isotropic when using the variogram model for kriging.

3.2.2 Isotropy/Anisotropy. Isotropy occurs when the range and sill remain unchanged as the functional variogram is calculated in different directions. If isotropy exists then the variogram function is the same for all orientations of **h**. However, if the range and sill do change with orientation, then the data is anisotropic. There are two types of anisotropy: geometric and zonal. An indication of geometric

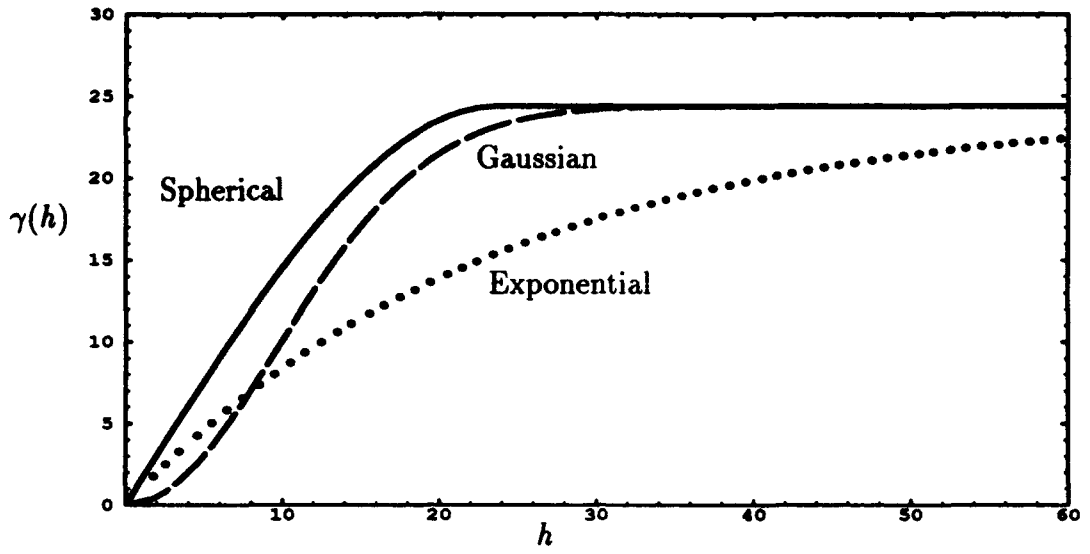


Figure 3.3 Functional Variogram Models: exponential, spherical, and Gaussian.

anisotropy is the change in value of a for two different directions with the sill remaining constant. Zonal isotropy occurs when the sill changes with direction (19). For geometric anisotropy a correction factor can be included in the final variogram model before kriging. Zonal anisotropy cannot be corrected with a simple correction factor.

The geometric anisotropic correction factor, k, is equal to the ratio of the a's, and always greater than or equal to 1 (when $k = 1$, then the data is isotropic):

$$k = \frac{a_{\text{maj}}}{a_{\text{min}}} \geq 1 \tag{3.13}$$

where a_{maj} and a_{min} denote the orientations of $\mathbf{h}$ with the largest and smallest range of influence (29). The correction factor can be incorporated into the functional variogram in a number of ways. One method is to include it directly into the functional variogram model (19). For example, if the spherical model is chosen, then the correction factor is used to scale h:

$$\gamma(h) = Co + C\left(\frac{3kh}{2a} - \frac{k^3h^3}{2a^3}\right) \tag{3.14}$$

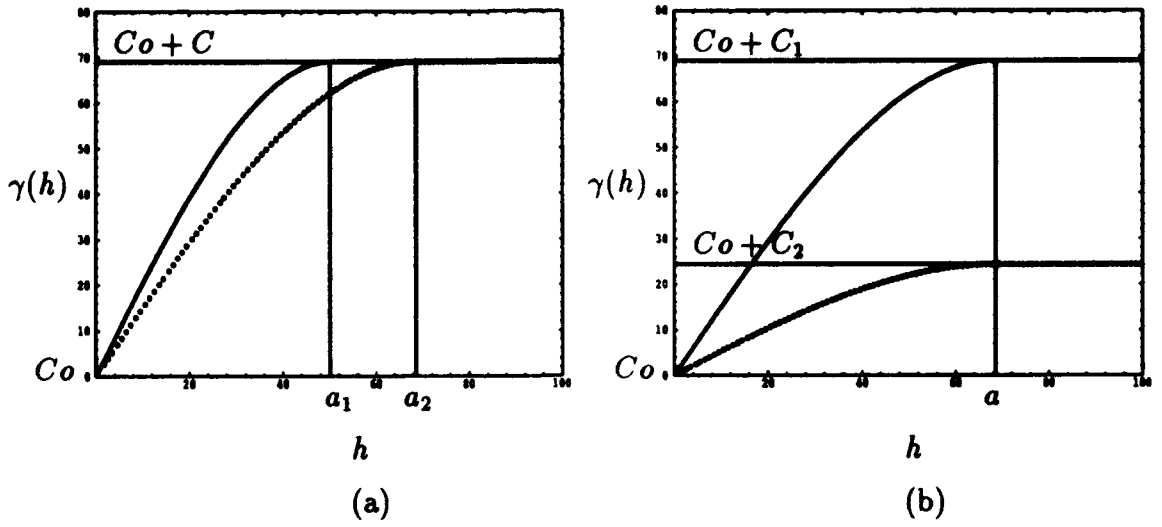


Figure 3.4 Variogram Anisotropies: (a)Geometric Anisotropy – range of influence changes with direction, (b) Zonal Anisotropy – sill changes with direction.

Another method is to incorporate the correction factor into the Euclidean distance calculation, before calculating the functional variogram (29):

$$h = \sqrt{(k_y(x_i - x_j))^2 + (k_x(y_i - y_j))^2} \tag{3.15}$$

where k_x and k_y are defined as:

$$\begin{aligned} &\text{If } a_x > a_y \quad \text{then: } a_x = a_{\text{maj}} \quad k_x = \tfrac{a_x}{a_y} \quad k_y = 1 \\ &\text{If } a_x < a_y \quad \text{then: } a_y = a_{\text{maj}} \quad k_x = 1 \quad k_y = \tfrac{a_y}{a_x} \end{aligned} \tag{3.16}$$

The variogram can also be used to indicate a trend in the data. If the data appear to have a trend, and violates the stationarity assumption, the trend must be removed before the variogram can be estimated. A common example of nonstationarity is a mean or expected value of the regionalized variable that changes with location. Trend removal is discussed in the next section.

3.2.3 Residual Analysis. The major assumption regarding the use of the variogram is that the data exhibit at least intrinsic stationarity. Intrinsic stationarity requires that there be no significant trend in the data. It is therefore generally desirable to remove the trend or bias from the data prior to kriging (19, 31).

The trend $z'(\mathbf{x})$ can often be approximated using a second order polynomial:

$$z'(\mathbf{x}) = b_1 + b_2 x + b_3 y + b_4 x^2 + b_5 y^2 + b_6 xy \tag{3.17}$$

The system of equations in matrix form, $\mathbf{z}' = \mathbf{Y}\mathbf{b}$, is:

$$\begin{bmatrix} z'_1 \\ z'_2 \\ z'_3 \\ z'_4 \\ \vdots \\ z'_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2 y_2 \\ 1 & x_3 & y_3 & x_3^2 & y_3^2 & x_3 y_3 \\ 1 & x_4 & y_4 & x_4^2 & y_4^2 & x_4 y_4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & x_n^2 & y_n^2 & x_n y_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \tag{3.18}$$

where n equals the total number of points in the data set. Since $\mathbf{z}$ and $\mathbf{Y}$ are known, the coefficients vector $\mathbf{b}$ can be estimated using least squares regression:

$$\mathbf{b} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{z} \tag{3.19}$$

The trend is subtracted from the measured $z(\mathbf{x})$ values to give the residuals $e(\mathbf{x})$:

$$e(\mathbf{x}) = z(\mathbf{x}) - z'(\mathbf{x}) \tag{3.20}$$

A functional variogram is found for the residual, and used in the kriging analysis. The resulting value from the kriging relationships is, in fact, an estimation of the residual at that location, i.e., $\hat{e}(\mathbf{x})$, and the desired value must include the drift

previously removed:

$$\hat{z}(\mathbf{x}) = \hat{e}(\mathbf{x}) + z'(\mathbf{x}) \tag{3.21}$$

where $\hat{z}(\mathbf{x})$ is an estimate of the regionalized variable at an unsampled location.

3.2.4 Kriging System. The kriging 'system' is the system of equations used to derive the BLUP, defined earlier as the best linear unbiased predictor for an unknown variable at a known location. There are several different types of kriging. This discussion only covers two: ordinary kriging and universal kriging. The difference between the two is treatment of any trend in the data. Complete knowledge of the trend allows the stationary variogram to be estimated and leads to ordinary kriging. Universal kriging assumes weak stationarity and incorporates the calculation of the trend in the final system of equations. The calculated trend is for the smaller or local neighborhood.

3.2.4.1 Ordinary Kriging. Ordinary kriging, also known as punctual kriging or point kriging, estimates the value of the regionalized variable, $\hat{z}(\mathbf{x}_0)$ at an unsampled but known location, $\mathbf{x}_0$, by taking the weighted linear combination of the measured values in the neighborhood of point $\mathbf{x}_0$:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n} w_i z(\mathbf{x}_i) \tag{3.22}$$

where $i = 1, \ldots, n$ and n is the number of total points in the neighborhood. To ensure unbiasedness, the weights are normalized:

$$\sum_{i=1}^{n} w_i = 1 \tag{3.23}$$

The difference between the actual value, $z(\mathbf{x}_0)$, which is not known, and the estimated value, $\hat{z}(\mathbf{x}_0)$, is the estimation error:

$$\epsilon(\mathbf{x}_0) = \hat{z}(\mathbf{x}_0) - z(\mathbf{x}_0) \tag{3.24}$$

and the error variance can be expressed as:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n}(\hat{z}(\mathbf{x}_0) - z(\mathbf{x}_0))^2}{n} \tag{3.25}$$

The minimized estimated kriging variance is found by:

$$\sigma_k^2 = \sum_{i=1}^{n} w_i \gamma_{i0} \tag{3.26}$$

where γ_{i0} is the variogram developed in structure analysis for distances from all measured locations to the unsampled location.

The objective of kriging is to attempt to determine the weights or estimates which minimize this variation error. The kriging system of equations is derived through partial differentiation in order to obtain the following relationships:

$$\begin{array}{ccccccccc}
w_1\gamma(h_{11}) & + & w_2\gamma(h_{12}) & + & \cdots & + & w_n\gamma(h_{1n}) & = & \gamma(h_{10}) \\
w_1\gamma(h_{21}) & + & w_2\gamma(h_{22}) & + & \cdots & + & w_n\gamma(h_{2n}) & = & \gamma(h_{20}) \\
\vdots & & & & & & & & \\
w_1\gamma(h_{n1}) & + & w_2\gamma(h_{n2}) & + & \cdots & + & w_n\gamma(h_{nn}) & = & \gamma(h_{n0}) \\
w_1 & + & w_2 & + & \cdots & + & w_n & = & 1.0
\end{array} \tag{3.27}$$

Since the number of unknowns is less than the number of equations, a slack variable λ, known as a Lagrangian multiplier is included. This adds an extra degree of

freedom to assure the solution has the minimum possible estimation error:

$$\begin{array}{ccccccccccc}
w_1\gamma(h_{11}) & + & w_2\gamma(h_{12}) & + & \cdots & + & w_n\gamma(h_{1n}) & + & \lambda & = & \gamma(h_{10}) \\
w_1\gamma(h_{21}) & + & w_2\gamma(h_{22}) & + & \cdots & + & w_n\gamma(h_{2n}) & + & \lambda & = & \gamma(h_{20}) \\
\vdots & & & & & & & & & & \\
w_1\gamma(h_{n1}) & + & w_2\gamma(h_{n2}) & + & \cdots & + & w_n\gamma(h_{nn}) & + & \lambda & = & \gamma(h_{n0}) \\
w_1 & + & w_2 & + & \cdots & + & w_n & + & 0 & = & 1.0
\end{array} \tag{3.28}$$

Rewriting these equations in matrix form gives the following result:

$$\begin{bmatrix}
\gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 \\
\gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 \\
1 & 1 & \cdots & 1 & 0
\end{bmatrix}
\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix}
=
\begin{bmatrix} \gamma(h_{10}) \\ \gamma(h_{20}) \\ \vdots \\ \gamma(h_{n0}) \\ 1 \end{bmatrix} \tag{3.29}$$

This equation can more easily be expressed as:

$$\mathbf{\Gamma w} = \gamma_0 \tag{3.30}$$

The optimum weights are calculated by:

$$\mathbf{w} = \mathbf{\Gamma}^{-1}\gamma_0 \tag{3.31}$$

Once the weights are derived, then the estimation $\hat{z}(\mathbf{x}_0)$ is calculated as:

$$\hat{z}(\mathbf{x}_0) = \mathbf{w}\mathbf{z}^{\mathrm{T}} \tag{3.32}$$

The minimized kriging variance is:

$$\sigma_k^2 = \mathbf{w}^{\mathrm{T}}\gamma_0 \tag{3.33}$$

which can be rewritten as:

$$\sigma_k^2(z_0) = \gamma_0^T \Gamma^{-1} \gamma_0 - ((1^T \Gamma^{-1} \gamma_0 - 1)^2 / (1^T \Gamma^{-1} 1)) \tag{3.34}$$

where **1** is a column vector of ones (18).

3.2.4.2 Universal Kriging. The residual analysis removes the global trend of the data; however, there may still be some localized trend. Universal kriging attempts to define simultaneously the optimum weights and the coefficients of this trend or drift in the local neighborhood. It is important to note that the size of the neighborhood impacts the smoothness of the drift, the size of the residuals, and the complexity of the variogram. Given a small localized neighborhood, the drift is less smooth, but the residuals are smaller and the variogram simpler. A larger neighborhood may give a smoother drift, but the residuals are larger, and the variogram more complex. The residual analysis discussion presented earlier could be applied to the entire data set, or a larger neighborhood, to provide a smoother drift. The residuals could then be used as the data in universal kriging to identify the BLUP and simultaneously to define the drift in the smaller, localized neighborhood. Assuming linear drift, the calculation of the localized residual is included in the kriging system, where the coefficients, a_1 and a_2 are similar to the coefficients, **b**, from the residual analysis discussion:

$$\begin{array}{ccccccccccccccc}
w_1\gamma(h_{11}) & + & w_2\gamma(h_{12}) & + & \cdots & + & w_n\gamma(h_{1n}) & + & \lambda & + & a_1x_1 & + & a_2y_1 & = & \gamma(h_{10}) \\
w_1\gamma(h_{21}) & + & w_2\gamma(h_{22}) & + & \cdots & + & w_n\gamma(h_{2n}) & + & \lambda & + & a_1x_2 & + & a_2y_2 & = & \gamma(h_{20}) \\
\vdots & & & & & & & & & & & & & & \\
w_1\gamma(h_{n1}) & + & w_2\gamma(h_{n2}) & + & \cdots & + & w_n\gamma(h_{nn}) & + & \lambda & + & a_1x_n & + & a_2y_n & = & \gamma(h_{n0}) \\
w_1 & + & w_2 & + & \cdots & + & w_n & + & 0 & + & 0 & + & 0 & = & 1.0 \\
w_1x_1 & + & w_2x_2 & + & \cdots & + & w_nx_n & + & 0 & + & 0 & + & 0 & = & x_0 \\
w_1y_1 & + & w_2y_2 & + & \cdots & + & w_ny_n & + & 0 & + & 0 & + & 0 & = & y_0
\end{array} \tag{3.35}$$

where x_0 and y_0 are the coordinates of the location where the height, $z_0(\mathbf{x})$ is being estimated. Rewriting these equations in matrix form:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 & x_1 & y_1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 & x_n & y_n \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \gamma(h_{10}) \\ \gamma(h_{20}) \\ \vdots \\ \gamma(h_{n0}) \\ 1 \\ x_0 \\ y_0 \end{bmatrix} \tag{3.36}$$

Note that these matrices are slightly different from those presented for ordinary kriging:

$$\boldsymbol{\Gamma}\mathbf{w} = \boldsymbol{\gamma}_0 \tag{3.37}$$

since universal kriging weights include the localized residual analysis, and the universal kriging $\boldsymbol{\gamma}_0$ includes the coordinates of the unsampled location. The rest of the calculations are similar to ordinary kriging:

$$\mathbf{w} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 \tag{3.38}$$

$$\hat{z}(\mathbf{x}_0) = \mathbf{w}^{\mathrm{T}}\mathbf{z} \tag{3.39}$$

where $\mathbf{z} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n), 0, 0, 0]^{\mathrm{T}}$ and $\mathbf{w} = [w_1, \ldots, w_n, \lambda, a_1, a_2]^{\mathrm{T}}$.

3.2.4.3 Dual-Kriging. Universal kriging can be reformulated into another form, known as the dual formation of kriging or dual-kriging (18, 31), where the weights are found from the following linear relationships:

$$\boldsymbol{\Gamma}_{\mathbf{n}}\hat{\mathbf{w}}_1 + \mathbf{X}\hat{\mathbf{w}}_2 = \mathbf{z} \tag{3.40}$$

$$\mathbf{X}^{\mathrm{T}}\hat{\mathbf{w}}_1 = \mathbf{0} \tag{3.41}$$

The submatrix, Γ_n, is the top left $n \times n$ elements of Γ, $\mathbf{0}$ is a vector of zeros, and:

$$\hat{\mathbf{w}}_1 = \begin{bmatrix} \hat{w}_1 \\ \vdots \\ \hat{w}_n \end{bmatrix} \quad \hat{\mathbf{w}}_2 = \begin{bmatrix} \hat{\lambda} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z(\mathbf{x}_1) \\ \vdots \\ z(\mathbf{x}_n) \end{bmatrix}$$

The new weights $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2]^T$ differ from the universal kriging weights (Eq. 3.38): $\hat{\mathbf{w}} \neq \mathbf{w}$. Expanding the dual-kriging equations (Eqs. 3.45, 3.41) shows how dual-kriging is equivalent to universal kriging:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_n \end{bmatrix} + \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \begin{bmatrix} \hat{\lambda} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} z(\mathbf{x}_1) \\ z(\mathbf{x}_2) \\ \vdots \\ z(\mathbf{x}_n) \end{bmatrix} \quad (3.42)$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (3.43)$$

where the residual analysis is partitioned out of the universal kriging Γ matrix and the vector of weights. These are combined into a form similar to universal kriging (Eq. 3.36):

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 & x_1 & y_1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 & x_n & y_n \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_n \\ \hat{\lambda} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} z(\mathbf{x}_1) \\ z(\mathbf{x}_2) \\ \vdots \\ z(\mathbf{x}_n) \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (3.44)$$

or:

$$\Gamma \hat{\mathbf{w}} = \mathbf{z} \quad (3.45)$$

Solving these equations for the new weights, $\hat{\mathbf{w}}$, results in:

$$\hat{\mathbf{w}} = \boldsymbol{\Gamma}^{-1}\mathbf{z} \tag{3.46}$$

where the weights in dual-kriging, $\hat{\mathbf{w}}$, are derived from known data, irrespective of the unsampled location. The universal kriging weights, $\mathbf{w}$ are dependent on the unsampled location, $\mathbf{x}_0$, through γ_0 (Eq. 3.38):

$$\mathbf{w} = \boldsymbol{\Gamma}^{-1}\gamma_0 \tag{3.47}$$

The BLUP equation is written as:

$$\hat{z}(\mathbf{x}_0) = \mathbf{w}^{\mathrm{T}}\mathbf{z} \tag{3.48}$$

$$= \gamma_0^{\mathrm{T}}\hat{\mathbf{w}} \tag{3.49}$$

Substituting in the definitions of the weights results in:

$$\hat{z}(\mathbf{x}_0) = \gamma_0^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{z} \tag{3.50}$$

Which is valid for either formulation. The equation in dual-kriging format is:

$$\hat{z}(\mathbf{x}_0) = \hat{\mathbf{w}}_1^{\mathrm{T}}\gamma_0 + \hat{\mathbf{w}}_2^{\mathrm{T}}\mathbf{x}_0 \tag{3.51}$$

$$= \sum_{i=1}^{n}\hat{w}_i\gamma(h_{i0}) + (\hat{\lambda} + \hat{a}_1 x_0 + \hat{a}_2 y_0) \tag{3.52}$$

Substituting in the value for the dual-kriging weights $\hat{\mathbf{w}}_1$, the BLUP equation can be written as:

$$\hat{z}(\mathbf{x}_0) = \gamma_0^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\mathbf{z} + (\hat{\lambda} + \hat{a}_1 x_0 + \hat{a}_2 y_0) \tag{3.53}$$

where the term in the parenthesis is the kriging regression.

Well	*Location,* $\mathbf{x}$	*Water Elevation,* $z(\mathbf{x})$
Well 1	(3.0,4.0)	120.0
Well 2	(6.3,3.4)	103.0
Well 3	(2.0,1.3)	142.0
Point $\mathbf{x}_0$	(3.0,3.0)	?

Table 3.1 Kriging Example: Water table elevation data.

3.2.5 Ordinary Kriging Example. An example is the best illustration of how kriging is applied, and the relative simplicity of the method. This example uses ordinary kriging to calculate the regionalized variable at an unsampled location. Recall that to do this, the data is assumed to be trend free and weakly stationary. An example previously described by Davis (20:pp. 386–392) is used where water table elevations are characterized based on three well samples. The water table depth at a new well with known coordinates $\mathbf{x}_0$ is to be predicted. Table 3.1 shows the well locations $\mathbf{x}$ and water table elevations $z(\mathbf{x})$. The variogram for this example is determined to be linear (i.e., a linearly increasing correlation), $\gamma(h) = Ah + Co$; with no nugget effect, $Co = 0$; and with slope $A = 4.0$ m^2/km. The distances between wells, h_{ij}, are calculated from the coordinates. These distances are mapped into $\gamma(h_{ij})$ by using the linear variogram model, $\gamma(h_{ij}) = 4.0h_{ij}$. In a similar manner, the well coordinate $\mathbf{x}_0$ is used to calculate h_{i0} and then $\gamma(h_{i0})$.

The ordinary kriging system of equations is applied to derive the unbiased weights, which are then used to calculate the BLUP, the best linear unbiased predictor, $\hat{z}(\mathbf{x}_0)$. The kriging equations are written as (Eq. 3.29):

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{12}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{10}) \\ \gamma(h_{20}) \\ \gamma(h_{30}) \\ 1 \end{bmatrix} \tag{3.54}$$

Substituting in the calculations for $\gamma(h_{ij})$ and $\gamma(h_{i0})$, this becomes:

$$\begin{bmatrix} 0 & 13.4 & 11.5 & 1 \\ 13.4 & 0 & 19.1 & 1 \\ 11.5 & 19.1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 4.0 \\ 13.3 \\ 7.9 \\ 1 \end{bmatrix} \tag{3.55}$$

The weights plus the Lagrangian slack variable are calculated using Equation 3.31:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.6039 \\ 0.0868 \\ 0.3093 \\ -0.7267 \end{bmatrix} \tag{3.56}$$

The best linear unbiased prediction of the water table height at point $\mathbf{x}_0$ is:

$$\begin{aligned} \hat{z}(\mathbf{x}_0) &= \textstyle\sum_{i=1}^{3} w_i z(\mathbf{x}_i) \\ &= 0.6039(120.0) + 0.0868(103.0) + 0.3093(142.0) \\ &= 125.3 \text{ meters} \end{aligned}$$

with a minimum error variance of:

$$\begin{aligned} \sigma_k^2 &= \mathbf{w}^{\mathrm{T}} \gamma_0 \\ &= 0.6039(4.0) + 0.0868(13.3) + 0.3093(7.9) - 0.7267 \\ &= 5.3 \text{ m}^2 \end{aligned}$$

Assuming a Gaussian distribution, the confidence interval of the prediction is calculated by:

$$\hat{z}(\mathbf{x}_0) \pm 1.96\sigma_k$$

which is the prediction for a 95 percent confidence interval. Applying the water table minimum variance using ordinary kriging methods gives:

$$\hat{z}(\mathbf{x_0}) = 125.3 \pm 4.5 \text{ m}$$

which is the prediction of the water table elevation at point $\mathbf{x_0}$ with 95 percent confidence.

3.2.6 Kriging Summary. There are several advantages to using kriging for form characterization. The main advantage is the combination of spacial correlation with minimum estimation variance. Also, by selecting a specific variogram model, the solution is guaranteed to be unique. Recall that kriging was developed as a geostatistic tool for obtaining optimum estimates of a regionalized variable at unsampled locations. Kriging has been applied in many other fields (6, 40), and even in the field of anthropology (27), but it has never been used for form comparison. The TPS method uses similar techniques for building the matrices which capture form description ($\mathbf{\Gamma} = \mathbf{L}$), but does not include the spacial correlation of the data. Therefore, kriging has the ability to describe uniquely the form of an object and can be used to derive relationships between similar objects.

3.3 Comparison of TPS and Kriging

Kriging and thin plate splines, TPSs, are closely related. The kriging and TPS equations to estimate $\hat{z}(x, y)$ are identical in form. However, kriging takes into account the actual spacial correlation of the data through a variogram function while TPS relates points through the minimum energy function $U(r) = r^2 \ln r^2$. The form characterization techniques of TPS can be greatly improved by incorporating the spacial correlation of the data.

3.3.1 Mathematical Relationship. The mathematical relationship between TPS and kriging can be seen by comparing the equations. Looking at the kriging equations first, the dual-kriging best linear unbiased predictor equation is (see Eq. 3.51):

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n} \hat{w}_{k_i} \gamma(h_{i0}) + (\hat{\lambda} + \hat{a}_1 x_0 + \hat{a}_2 y_0) \tag{3.57}$$

where $\gamma(h)$ is the variogram function found through structure analysis to best represent the spacial correlation of the data. The dual-kriging weights, $\hat{\mathbf{w}}_k$, are calculated by (see Section 3.2.4.3, Eqs. 3.44, 3.45, 3.46):

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 & x_1 & y_1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 & x_n & y_n \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_{k_1} \\ \hat{w}_{k_2} \\ \vdots \\ \hat{w}_{k_n} \\ \hat{\lambda} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} z(\mathbf{x}_1) \\ z(\mathbf{x}_2) \\ \vdots \\ z(\mathbf{x}_n) \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.58}$$

$$\boldsymbol{\Gamma} \hat{\mathbf{w}}_k = \mathbf{z} \tag{3.59}$$

$$\hat{\mathbf{w}}_k = \boldsymbol{\Gamma}^{-1} \mathbf{z} \tag{3.60}$$

where $\hat{\mathbf{w}}_k = [\hat{w}_{k_1}, \ldots, \hat{w}_{k_n}, \hat{\lambda}, \hat{a}_1, \hat{a}_2]^{\mathrm{T}}$, and $\boldsymbol{\Gamma}^{-1}$ is a symmetric matrix.

From the thin plate spline equations, the predicted height of a landmark (see Eq. 2.18) is:

$$\hat{z}(x_0, y_0) = \sum_{i=1}^{n} U(r_{i0}) \hat{w}_{t_i} + c_1 + c_2 x_0 + c_3 y_0 \tag{3.61}$$

where $U(r_{ij})$ is the minimized bending energy function. The TPS weights are calculated (see Eq. 2.19) by:

$$\begin{bmatrix} U(r_{11}) & U(r_{12}) & \dots & U(r_{1n}) & 1 & x_1 & y_1 \\ U(r_{21}) & U(r_{22}) & \dots & U(r_{2n}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ U(r_{n1}) & U(r_{n2}) & \dots & U(r_{nn}) & 1 & x_n & y_n \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 \\ x_1 & x_2 & \dots & x_n & 0 & 0 & 0 \\ y_1 & y_2 & \dots & y_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_n \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.62}$$

$$\mathbf{L}\hat{\mathbf{w}}_t = \mathbf{z} \tag{3.63}$$

$$\hat{\mathbf{w}}_t = \mathbf{L}^{-1}\mathbf{z} \tag{3.64}$$

where $\hat{\mathbf{w}}_t = [\hat{w}_{t_1}, \dots, \hat{w}_{t_n}, c_1, c_2, c_3]^{\mathrm{T}}$ and $\mathbf{L}^{-1}$ is a symmetric matrix.

Comparing Equations 3.60 and 3.64, the TPS and kriging weights are derived the same way. Similarly, the TPS and kriging predictors (Eqs. 3.57, 3.61) are derived in identical formulations except for the important fact that $\gamma(h_{ij}) \neq U(r_{ij})$, because $\gamma(h_{ij})$ represents the actual spacial correlation of the data, and $U(r_{ij})$ is the minimum energy function. Therefore, the thin plate spline and kriging equations are equivalent except for the variogram function. Example 5 (Sec. 2.2.4.10) uses TPS to compare the form of two objects. The results of comparing objects can be greatly improved if the actual spacial correlation of the data is considered.

3.3.2 Error in Development of TPS. According to Bookstein (4), the thin plate spline minimizes the total curvature (Eq. 2.5):

$$\oint\!\!\oint_{\Re^2} \left[\left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 \right] dx dy \tag{3.65}$$

However, as noted by Cressie (18:p. 181), the TPS is *not* the minimization of this integral.

From Cressie, the two-dimensional Laplacian smoothing spline of degree 2 is:

$$\hat{z}(\mathbf{s}_0) = \sum_{i=1}^{n} b_i e(\mathbf{s}_0 - \mathbf{s}_i) + a_0 + a_1 x_0 + a_2 y_0 \tag{3.66}$$

where $\mathbf{s}_0 = [x_0, y_0]^{\mathrm{T}}$, $e(\mathbf{h}) = \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|^2)/(16\pi)$, $\mathbf{a} = [a_0, a_1, a_2]^{\mathrm{T}}$, and $\mathbf{b} = [b_0, \ldots, b_n]^{\mathrm{T}}$. By comparison, this equation is equivalent to the TPS predictor (Eq. 3.61). The Laplacian spline constants **a** and **b** are found by simultaneously solving the $n + 3$ linear equations:

$$(\mathbf{K} + n\rho\mathbf{I})\mathbf{b} + \mathbf{X}\mathbf{a} = \mathbf{z} \tag{3.67}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{b} = \mathbf{0} \tag{3.68}$$

where $n\rho$ is the nugget effect (Co) or the measurement error, **K** is the upper $n \times n$ elements of **L**, and:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \tag{3.69}$$

The TPS solution for the constants is given in Equation 3.64. Equation 3.67 is equivalent to the first n equations of Equation 3.62 for $\rho = 0$. Equation 3.68 is equivalent to the last three equations of Equation 3.61.

Therefore, the TPS is a two-dimensional Laplacian smoothing spline of degree 2 with $\rho = 0$. According to Cressie, a spline of this type minimizes:

$$\sum_{i=1}^{n} [z(\mathbf{s}_i) - g(\mathbf{z}; \mathbf{s}_i)]^2 / n + \rho \iint_{\Re^2} \left[\left(\frac{\partial^2 g}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 g}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 g}{\partial y^2} \right)^2 \right] dx dy \tag{3.70}$$

where g is the smoothing function equivalent to $U(r)$. Since $\rho = 0$, the thin plate spline actually minimizes:

$$\sum_{i=1}^{n} [z(\mathbf{s}_i) - g(\mathbf{z}; \mathbf{s}_i)]^2 / n \tag{3.71}$$

so Bookstein's statement that the TPS minimizes the double integral is incorrect. However, Bookstein's final spline equations are useful.

3.4 Energy Matrix

The form characterization method is developed as a combination of kriging and TPS. This results in a structure analysis technique which provides spacial correlation of the data points merged with the bending energy definition from the TPS method. The form characterization method creates a bending energy matrix or energy matrix for each face. This energy matrix is significant because it captures the unique form of the face for comparison. By inserting the variogram model into the universal kriging matrix, the following **L** matrix develops:

$$\mathbf{L} = \begin{bmatrix} \nwarrow & \uparrow & \nearrow & 1 & x_1 & y_1 \\ \leftarrow & \mathbf{K} & \rightarrow & \vdots & \vdots & \vdots \\ \swarrow & \downarrow & \searrow & 1 & x_n & y_n \\ 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \tag{3.72}$$

where:

$$\mathbf{K} = \begin{bmatrix} 0 & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) \\ \gamma(h_{21}) & 0 & \cdots & \gamma(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & 0 \end{bmatrix} \tag{3.73}$$

The variogram, $\gamma(h_{ij})$, is a function of the scalar distance h_{ij} from (x_i, y_i) to (x_j, y_j). Two seal areas with similar variances have different energy matrices since the separation distance h_{ij} varies from face to face, depending on size and shape.

The $\mathbf{L}$ matrix contains all the available information about the form of a face within the seal area. It is the spacial correlation plus a localized residual analysis. In the TPS method, $\mathbf{L}^{-1}$ is referred to as the energy matrix (5). The $\mathbf{L}^{-1}$ in kriging presents a unique description of a face: The Best Linear Unbiased Predictor (BLUP) (see Section 3.2). The BLUP for z at the point (x_0, y_0) is given by (Eq. 3.50):

$$\hat{z}_0 = \gamma_0^{\mathrm{T}} \mathbf{L}^{-1} \mathbf{z} \tag{3.74}$$

where $\mathbf{z}$ is a vector of known measurements, and γ_0 is a vector corresponding to h_{i0}, the distances from all other locations to (x_0, y_0): $\mathbf{z} = [z_1, \ldots, z_n, 0, 0, 0]^{\mathrm{T}}$ and $\gamma_{i0} = [\gamma_{10}, \ldots, \gamma_{n0}, 1, x_0, y_0]^{\mathrm{T}}$.

In the BLUP equation (Eq. 3.74), $\mathbf{z}$ forces the points to conform exactly to the data and γ_0 accounts for the specific location where the regionalized variable, z, is being predicted. The remaining term, $\mathbf{L}^{-1}$, describes the overall shape or bending of the surface passing through the data. Therefore, $\mathbf{L}^{-1}$ is the energy matrix describing the shape of each mask seal area. The last three rows and columns of $\mathbf{L}^{-1}$ may be ignored in the derivation because they represent the residual analysis and an affine transformation. Therefore, the top $n \times n$ elements of $\mathbf{L}^{-1}$, denoted by $\mathbf{L}_n^{-1}$ is typically referred to as the energy matrix.

The energy matrix is created without the use of landmarks. Therefore, each face contains a different quantity of data dependent on their relative size. That is, the total number of data points, n, is not constant from face to face because small faces have less data than large faces. This results in different energy matrix sizes. The energy matrix is the source for clustering. In order to conduct a comparison of seal areas, the identification of a set of minimum unique features is required. The

number of features can be varied, but must be uniform from face to face for input to the clustering routine. One method of feature space determination is to identify the unique features per individual face from the uneven data sets. This is called the 'free form' method. This section describes an application of the free form method which uses the eigensystem of each energy matrix to create a reduced feature space for clustering.

3.4.1 Energy Matrix Decomposition. The energy matrix must be decomposed into a small number of features for clustering. Spectral decomposition and singular value decomposition are two techniques to decompose the energy matrix using eigenvalues and eigenvectors. The eigenvalues could then, potentially, be used as features for clustering.

Since the $\mathbf{L}$ matrix is symmetric, the energy matrix $\mathbf{L}_n^{-1}$ is also symmetric. Spectral decomposition (also called the principal axes theorem) describes the factorization of a real symmetric matrix into:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}} \tag{3.75}$$

where $\mathbf{\Lambda}$ is a diagonal matrix consisting of the eigenvalues (λs) of $\mathbf{A}$, and $\mathbf{Q}$ contains the corresponding orthonormal eigenvectors (50). Using spectral decomposition, $\mathbf{L}_n^{-1}$ can be decomposed into its eigensystem, which is unique to each matrix. The eigenvalues then present a possible feature space for comparison, reducing the amount of data from n^2 to n.

The singular value decomposition (SVD) for any matrix is:

$$\mathbf{A} = \mathbf{Q}_1\mathbf{\Sigma}\mathbf{Q}_2^{\mathrm{T}} \tag{3.76}$$

or:

$$\mathbf{Q}_1^{\mathrm{T}}\mathbf{A}\mathbf{Q}_2 = \mathbf{\Sigma} \tag{3.77}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix. The columns of $\mathbf{Q}_1$ are the eigenvectors of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$, the columns of $\mathbf{Q}_2$ are the eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, and the singular values on the diagonal of $\mathbf{\Sigma}$ are the square roots of the nonzero eigenvalues of both $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ (50). If $\mathbf{A}$ is symmetric (as in $\mathbf{L}^{-1}$), then $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$. This means that the diagonal elements of $\mathbf{\Sigma}$ are equal to $\sqrt{\lambda^2}$. For real symmetric matrices, SVD is identical to the spectral decomposition (49) regardless of the sign of the eigenvalues. Therefore, the magnitude of the eigenvalues can be used as features for clustering.

As described by Bookstein, the eigensystem of the energy matrix, $\mathbf{L}_n^{-1}$ uniquely captures the form of a face (5:p. 323). The eigenvalues represent the magnitude of the bending energy, and the related eigenvectors identify the location of the bending. The next section investigates bending modes to determine the bending associated with an individual eigenvalue eigenvector combination. The use of bending modes is also explored to verify the ability to reconstruct an object from the eigensystem using the spectral theorem.

3.4.2 Bending Modes. Bending modes utilize the eigensystem to decompose the bending energy matrix or energy matrix. Each bending mode represents a surface plot of an individual eigenvalue-eigenvector combination. The bending modes are compared to show the importance of each eigenvalue-eigenvector combination.

3.4.2.1 Development of Bending Modes. Recall that the dual-kriging estimation of the height of a face at the point (x_0, y_0) is:

$$\hat{z}(x_0, y_0) = \gamma_0^{\mathrm{T}}\mathbf{L}_n^{-1}\mathbf{z} + (c_1 + c_2 x_0 + c_3 y_0) \tag{3.78}$$

where $\gamma_0 = [\gamma_{01}, \ldots, \gamma_{0n}]^{\mathrm{T}}$ and $\mathbf{z} = [z_1, \ldots, z_n]^{\mathrm{T}}$. The point estimation is an interpolation of the face between spacial delineators. The two parts of the interpolation are

the bending ($\gamma^{\mathrm{T}}\mathbf{L}_n^{-1}\mathbf{z}$) and the regression ($c_1 + c_2x_0 + c_3y_0$). The regression constants are deterministic as shown in Section 3.2.4.2.

The energy matrix is a real symmetric matrix so it can be spectrally decomposed into:

$$\mathbf{L}_n^{-1} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i \mathbf{r}_i{}^{\mathrm{T}} \tag{3.79}$$

where λ, $\mathbf{e}$, and $\mathbf{r}$ are the eigenvalues, right eigenvectors, and left eigenvectors of the bending energy matrix, respectively (50). The left eigenvectors are related to the right eigenvectors by:

$$\mathbf{R} = [\mathbf{r}_1|\mathbf{r}_2|\cdots|\mathbf{r}_n] = \left[[\mathbf{e}_1|\mathbf{e}_2|\cdots|\mathbf{e}_n]^{-1}\right]^{\mathrm{T}} \tag{3.80}$$

Substituting spectral decomposition (Eq. 3.79) into the interpolation (Eq. 3.78) gives:

$$z(x_0, y_0) = \gamma_0^{\mathrm{T}} \left[\sum_{i=1}^{n} \lambda_i \mathbf{e}_i \mathbf{r}_i^{\mathrm{T}}\right] \mathbf{z} + (c_1 + c_2x_0 + c_3y_0) \tag{3.81}$$

$$= \sum_{i=1}^{n} \Omega_i + (c_1 + c_2x_0 + c_3y_0) \tag{3.82}$$

where Ω_i is a scalar defined to be the i^{th} bending mode of the face:

$$\Omega_i = \Omega_i(x_0, y_0) = \lambda_i \gamma_0^{\mathrm{T}} \mathbf{e}_i \mathbf{r}_i{}^{\mathrm{T}} \mathbf{z} \tag{3.83}$$

If the eigenvalues are unique, then the left eigenvectors are identical to the right eigenvectors (49). However, for the energy matrix, there are always three repeated zero eigenvalues representing the affine transformation of the regression plane. In all test cases performed for this research, the right and left eigenvectors associated with a non-zero eigenvalue are identical; therefore $\mathbf{e}_i$ can be substituted for $\mathbf{r}_i$. For the three zero eigenvalues, the right and left eigenvectors are different, but the bending mode is *always* zero because $\lambda_i = 0$. Therefore, the bending modes can be simplified

to:

$$\Omega_i = \Omega_i(x_0, y_0) = \lambda_i \gamma_0^{\mathrm{T}} \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}} \mathbf{z} \tag{3.84}$$

The bending mode is a scalar function of (x_0, y_0). Each bending mode consists of three scalars: λ_i is the i^{th} eigenvalue, $\psi_i = \mathbf{e}_i^{\mathrm{T}}\mathbf{z}$ is a deterministic constant called the mode constant, and $\gamma_0^{\mathrm{T}}\mathbf{e}_i$ is a function of (x_0, y_0) representing the shape of the bending mode.

3.4.2.2 Bending Mode Example. This example explains the procedure for decomposing the energy matrix into individual bending modes. This example is a continuation of Example 3, Section 2.2.4.7. For simplicity, the same data is used with variogram function $\gamma(h) = h^2 \ln h^2$.

The five data points describing the figure are $(2,4,7)$, $(3,10,2)$, $(-4,7,1)$, $(7,-2,-4)$, and $(4,1,5)$. The variogram function is $\gamma(h_{i0}) = h_{i0}^2 \ln h_{i0}^2$ where $h_{i0}^2 = [(x_i - x_0)^2 + (y_i - y_0)^2]$ is the distance squared from landmark i to the point (x_0, y_0). This variogram function is not representative of a face; however, it is used here for simplification. These results hold for any variogram function.

In the previous example, the interpolation of the data is found to be:

$$\begin{aligned} z(x,y) \;=\; & 0.0326\,\gamma\left(\sqrt{(x-2)^2+(y-4)^2}\right) - 0.0101\,\gamma\left(\sqrt{(x-3)^2+(y-10)^2}\right) \\ & -0.0240\,\gamma\left(\sqrt{(x+4)^2+(y-7)^2}\right) - 0.0457\,\gamma\left(\sqrt{(x-7)^2+(y+2)^2}\right) \\ & +0.0471\,\gamma\left(\sqrt{(x-4)^2+(y-1)^2}\right) + 22.121 - 0.3381x + 0.2234y \end{aligned} \tag{3.85}$$

where the subscript is dropped since (x_0, y_0) can be any location. The regression of the data is given by the last three terms:

$$\text{Regression} = 22.121 - 0.3381x + 0.2234y \tag{3.86}$$

The **L** matrix is:

$$\mathbf{L} = \begin{bmatrix} 0 & \gamma(\sqrt{37}) & \gamma(\sqrt{45}) & \gamma(\sqrt{61}) & \gamma(\sqrt{13}) & 1 & 2 & 4 \\ \gamma(\sqrt{37}) & 0 & \gamma(\sqrt{58}) & \gamma(\sqrt{160}) & \gamma(\sqrt{82}) & 1 & 3 & 10 \\ \gamma(\sqrt{45}) & \gamma(\sqrt{58}) & 0 & \gamma(\sqrt{202}) & \gamma(\sqrt{100}) & 1 & -4 & 7 \\ \gamma(\sqrt{61}) & \gamma(\sqrt{160}) & \gamma(\sqrt{202}) & 0 & \gamma(\sqrt{18}) & 1 & 7 & -2 \\ \gamma(\sqrt{13}) & \gamma(\sqrt{82}) & \gamma(\sqrt{100}) & \gamma(\sqrt{18}) & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & -4 & 7 & 4 & 0 & 0 & 0 \\ 4 & 10 & 7 & -2 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.87}$$

Substituting $\gamma(h) = h^2 \ln h^2$ into **L** and taking the inverse:

$$\mathbf{L}^{-1} = \begin{bmatrix} 0.02057 & -0.00341 & -0.00337 & 0.00358 & -0.01736 & 1.89040 & 0.01951 & -0.05364 \\ -0.00341 & 0.00060 & 0.00071 & -0.00015 & 0.00224 & -0.92498 & 0.09310 & 0.12493 \\ -0.00337 & 0.00071 & 0.00118 & 0.00115 & 0.00031 & -0.25472 & -0.11780 & -0.02708 \\ 0.00358 & -0.00015 & 0.00115 & 0.00540 & -0.00998 & -1.06067 & 0.06322 & -0.06635 \\ -0.01736 & 0.00224 & 0.00031 & -0.00998 & 0.02478 & 1.34998 & -0.05803 & 0.02215 \\ 1.89040 & -0.92498 & -0.25472 & -1.06067 & 1.34998 & 552.028 & -19.2475 & -31.3383 \\ 0.01951 & 0.09310 & -0.11780 & 0.06322 & -0.05803 & -19.2475 & 9.32795 & -1.39663 \\ -0.05364 & 0.12493 & -0.02708 & -0.06635 & 0.02215 & -31.3383 & -1.39663 & 9.49502 \end{bmatrix} \tag{3.88}$$

The upper left 5×5 elements of the $\mathbf{L}^{-1}$ is the energy matrix $\mathbf{L}_5^{-1}$:

$$\mathbf{L}_5^{-1} = \begin{bmatrix} 0.02057 & -0.00341 & -0.00337 & 0.00358 & -0.01736 \\ -0.00341 & 0.00060 & 0.00071 & -0.00015 & 0.00224 \\ -0.00337 & 0.00071 & 0.00118 & 0.00115 & 0.00031 \\ 0.00358 & -0.00015 & 0.00115 & 0.00540 & -0.00998 \\ -0.01736 & 0.00224 & 0.00031 & -0.00998 & 0.02478 \end{bmatrix} \tag{3.89}$$

The nonzero eigenvalues and eigenvectors of $\mathbf{L}_n^{-1}$ are:

$$\lambda_1 = 0.04324 \qquad \mathbf{e}_1 = [\,0.62274 \quad -0.09036 \quad -0.05010 \quad 0.25142 \quad -0.73369\,]^{\mathrm{T}}$$

$$\lambda_2 = 0.00931 \qquad \mathbf{e}_2 = [\,-0.63886 \quad 0.16488 \quad 0.34010 \quad 0.53599 \quad -0.40211\,]^{\mathrm{T}}$$

The mode constants are given by the dot product:

$$\begin{aligned}
\psi_1 &= \mathbf{e}_1{}^{\mathrm{T}}\mathbf{z} \\
&= [0.62274 \quad -0.09036 \quad -0.05010 \quad 0.25142 \quad -0.73369] \cdot \\
&\quad [7 \quad 2 \quad 1 \quad -4 \quad 5]^{\mathrm{T}} \\
&= -0.54577 \\
\psi_2 &= \mathbf{e}_2{}^{\mathrm{T}}\mathbf{z} \\
&= [-0.63886 \quad 0.16488 \quad 0.3401[illegible] \quad 0.53599 \quad -0.40211] \cdot \\
&\quad [7 \quad 2 \quad 1 \quad -4 \quad 5]^{\mathrm{T}} \\
&= -7.95667
\end{aligned}$$

The mode constants are deterministic scalars. The bending modes are given by:

$$\begin{aligned}
\Omega_1 &= \lambda_1 \gamma^{\mathrm{T}} \mathbf{e}_1 \psi_1 = \lambda_1 \psi_1 \gamma^{\mathrm{T}} \mathbf{e}_1 \\
&= (0.04324)(-0.54577) \times \\
&\quad \Big[0.62274\,\gamma\left(\sqrt{(x-2)^2+(y-4)^2}\right) - 0.09036\,\gamma\left(\sqrt{(x-3)^2+(y-10)^2}\right) \\
&\quad -0.05010\,\gamma\left(\sqrt{(x+4)^2+(y-7)^2}\right) - 0.25142\,\gamma\left(\sqrt{(x-7)^2+(y+2)^2}\right) \\
&\quad -0.73369\,\gamma\left(\sqrt{(x-4)^2+(y-1)^2}\right)\Big] \qquad (3.90) \\
\Omega_2 &= \lambda_2 \gamma^{\mathrm{T}} \mathbf{e}_2 \psi_2 = \lambda_2 \psi_2 \gamma^{\mathrm{T}} \mathbf{e}_2 \\
&= (0.00931)(-7.95667) \times \\
&\quad \Big[-0.63886\,\gamma\left(\sqrt{(x-2)^2+(y-4)^2}\right) + 0.16488\,\gamma\left(\sqrt{(x-3)^2+(y-10)^2}\right) \\
&\quad +0.34010\,\gamma\left(\sqrt{(x+4)^2+(y-7)^2}\right) + 0.53599\,\gamma\left(\sqrt{(x-7)^2+(y+2)^2}\right) \\
&\quad -0.40211\,\gamma\left(\sqrt{(x-4)^2+(y-1)^2}\right)\Big] \qquad (3.91)
\end{aligned}$$

The bending modes are only a function of x and y. The surface plots of these modes as x and y vary are given in Figures 3.5 and 3.6. The first bending mode is more 'bent' than the second bending mode. This shows that the bending of the first mode is more localized. The first bending mode relates to the higher eigenvalue. Therefore,

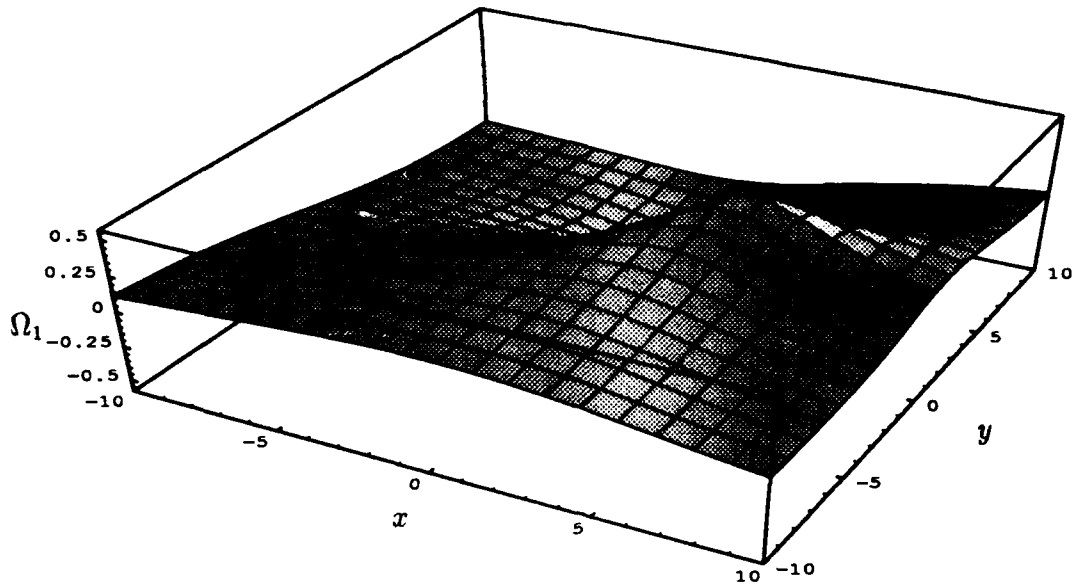


Figure 3.5 Example Bending Mode 1: Bending mode plus regression plane for five point example.

higher eigenvalues represent localized bending of the figure while lower eigenvalues represent more global changes in the figure.

The next step is to confirm that the bending modes can be combined into the original figure. For the landmark at (2,4) the bending modes are given by Equations 3.90 and 3.91. The regression is given by Equation 3.86. Substituting the landmark $(2,4)$ into these equations gives:

$$\Omega_1(2,4) = -0.42307 \tag{3.92}$$

$$\Omega_2(2,4) = -14.9152 \tag{3.93}$$

$$\text{Regression}(2,4) = 22.3383 \tag{3.94}$$

Adding the bending modes and regression (Eq. 3.81) gives $-0.42307 - 14.9152 + 22.338 = 7.0$ which is the height of the landmark at $(2,4)$. Similarly, the bending modes sum to the landmark at each of the five landmarks. The bending modes can also be calculated at non-landmark positions. For example, at the point $(5,3)$: $\Omega_1(5,3) = 0.19503$, $\Omega_2(5,3) = -16.3025$, and the regression is 21.1006. Adding these

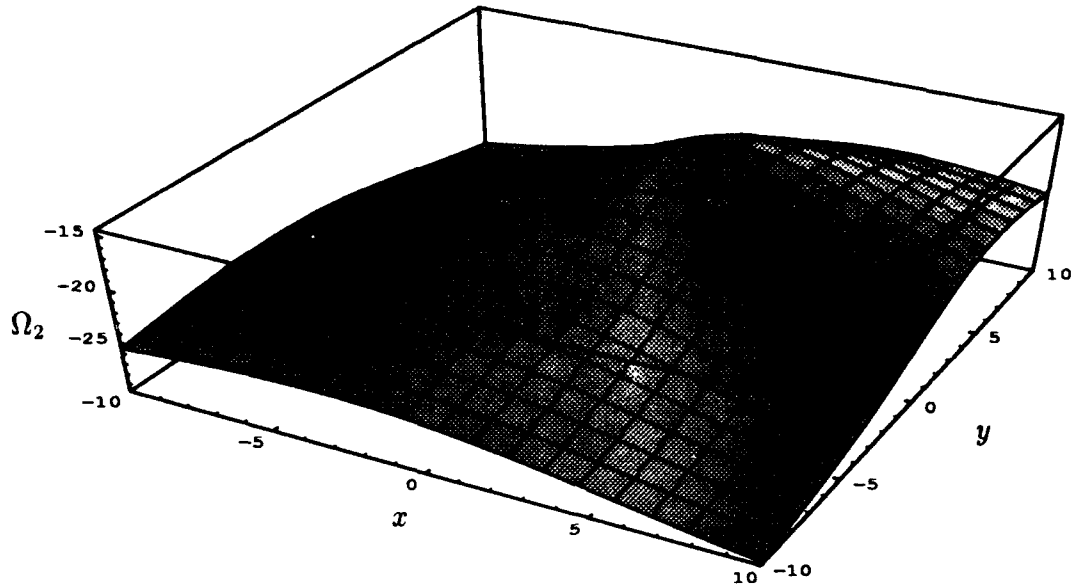


Figure 3.6 Example Bending Mode 2: Bending mode plus regression plane for five point example.

gives 4.9931. The interpolation from Equation 3.85 is also 4.9931 which shows the bending modes always recombine into the original figure.

This example has shown that the energy matrix can be spectrally decomposed into a series of bending modes and that the bending modes can be recombined into the original figure.

3.4.2.3 Bending Modes Summary. The eigenvalues, eigenvectors, and data points represent a complete picture of the energy matrix and of the face. However, the feature space used for clustering should be significantly less than the n^2 elements of the energy matrix. One way to reduce the size of the feature space is to use only the eigenvalues for clustering. As a result of this data reduction, the eigenvector and mode constant information is lost.

In order to cluster on eigenvalues alone, the relative importance of each eigenvalue must be determined. Bending modes are a valuable tool for 'viewing' the eigenvalues of a face. They can also show which eigenvalues are the important eigenvalues of a particular face. However, they can not be used to find, in general, which

eigenvalues are the important eigenvalues for comparison of facial seal data. Therefore, another technique must be used to infer which eigenvalues best represent the similarities and differences in facial seal data.

3.4.3 Investigation of Eigenvalues. The eigenvalues of the energy matrix, $\mathbf{L}_n^{-1}$, are unique. Eigenvalues alone represent a reduced feature space as compared to $\mathbf{L}_n^{-1}$ (n versus n^2 features), but every time a feature space is reduced, information is lost. If the eigenvalues are analyzed as a feature space, an investigation is needed to determine if they hold enough information to characterize the seal area. The next section validates use of the energy matrix eigenvalues as a feature space which captures changes in form. The validation uses a symmetric flat ellipse, or 'egg', to test the sensitivity of the eigenvalues to changes in shape and size.

3.4.3.1 The Egg. The 'egg' is developed to investigate whether or not the eigenvalues of $\mathbf{L}_n^{-1}$ contain enough information to characterize the seal area of a face. The egg consists of 18 (x, y, z) coordinate points. Figure 3.7 shows a plot of these points and Table 3.2 lists the (x, y, z) coordinates. The first 16 points define the circumference of an ellipse in the xy plane (height $z = 0$), symmetric about the x and y axes. The major (x-axis) and minor (y-axis) axes represent the length and width, respectively, of a typical seal area. The last two of the 18 points represent a nose and chin. They are placed on the x-axis (length of seal area) and given z coordinates typical of nose and chin measurements after trend removal of the facial seal areas (refer to the residual analysis used for trend removal in determining the variogram, Section 3.2.3).

The simplicity of the egg enables slight changes in form to be captured in $\mathbf{L}_n^{-1}$. These slight changes are made by modifying the egg described previously to create many different eggs for comparison. Modifications, changes in form of the egg, are made in several different ways. They are listed here and given an identifying name:

Elliptical Points	x	y	z
1	58.0	0.0000	0
2	43.5	29.1033	0
3	29.0	38.1051	0
4	14.5	42.6028	0
5	0.0	44.0000	0
6	-14.5	42.6028	0
7	-29.0	38.1051	0
8	-43.5	29.1033	0
9	-58.0	0.0000	0
10	-43.5	-29.1033	0
11	-29.0	-38.1051	0
12	-14.5	-42.6028	0
13	0.0	-44.0000	0
14	14.5	-42.6028	0
15	29.0	-38.1051	0
16	43.5	-29.1033	0
Nose	20.0	0	20.0
Chin	-40.0	0	10.0

Table 3.2 Original Egg Coordinate Data

- The egg is lengthened 30 % by scaling the x coordinates. (Egg-x130)
- The egg is widened 30 % by scaling the y coordinates. (Egg-y130)
- The egg is increased 30 % in both the x and y directions. (Egg-xy130)
- The egg is increased 30% in the x, y and z directions, making it the same shape, but a different size. (Egg-130)
- The height of the nose is changed by 5 units in both the positive and negative z direction creating two modified eggs. (Nose-up, Nose-down)
- The height of the chin is changed by 5 units in both the positive and negative z direction creating two modified eggs. (Chin-up, Chin-down)
- The position of the nose is moved along the x-axis by 5 units in both the positive and negative x direction creating two modified eggs. (Nose-xpos, Nose-xneg)

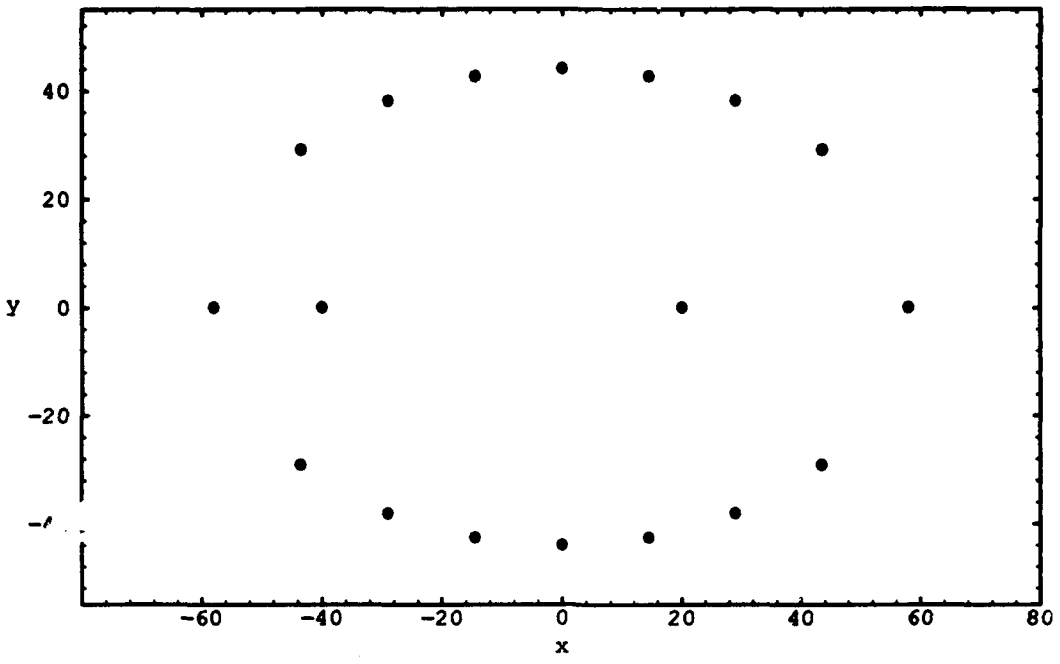
- The position of the chin is moved along the x-axis by 5 units in both the positive and negative x direction creating two modified eggs. (Chin-xpos, Chin-xneg)

The original egg and all modified eggs are analyzed using the structure analysis methods outlined in Section 3.2, with the exception of the residual analysis. The energy matrices for the egg and all modifications are developed using the spherical variogram model, with an anisotropic correction factor similar to what is used in the seal area data analyses. After the energy matrix, $\mathbf{L}_n^{-1}$, is calculated, the eigenvalues of this matrix are found and compared. The comparisons are done in three different groups: scaling effects (Egg-x130, Egg-y130, Egg-xy130, Egg-130), movement in height of the nose and chin (Nose-up, Nose-down, Chin-up, Chin-down), and movement of nose and chin along the x-axis (Nose-xpos, Nose-xneg, Chin-xpos, Chin-xneg). Comparisons are discussed with reference to the absolute value or magnitude of the eigenvalues, since singular value decomposition (Section 3.4.1) shows that the sign of eigenvalues for real symmetric matrices is not relevant for comparison.

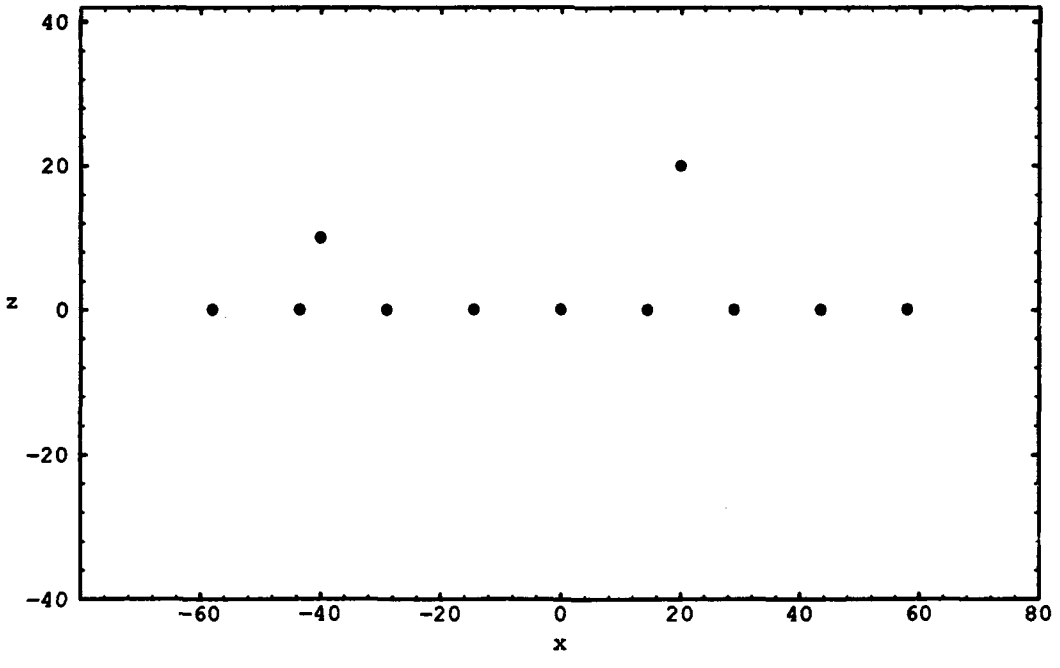
3.4.3.2 Scaling Effects. Five files are used in the scaling effects: the original egg, Egg-x130, Egg-y130, Egg-xy130, Egg-130. When comparing the eigenvalues of the energy matrix from these five eggs, a trend appears. Figure 3.8 shows that the relative magnitudes of the eigenvalues from the corresponding energy matrices follow the trend:

$$\text{original Egg} > \text{Egg-y130} \gg \text{Egg-x130} > \text{Egg-xy130} \gg \text{Egg-130}$$

The symbol $\gg$ represents a noticeably larger increase in eigenvalues, such as between Egg-x130 and Egg-y130, compared to differences represented by the symbol $>$. The trend shows that smaller sizes, the smallest represented by the original egg, have higher bending energies than larger sizes. Figure 3.8 also shows that the rate of decrease in the eigenvalues tends to level out around the eighth eigenvalue after which the lower eigenvalues of Egg-x130 and Egg-xy130 become slightly larger than the original egg and Egg-y130. Egg-y130 is smaller than Egg-x130 because increasing



(a)



(b)

Figure 3.7 Plot of the Egg: The original egg which consists of 18 points; (a) egg top view, (b) egg side view.

the width by 30% results in a smaller increase in size compared to increasing the length of the egg by 30%. Bending a plane through a small seal area requires more energy than bending a plane through a larger area and is shown here by the eigenvalues of the four eggs compared.

3.4.3.3 Movement in Height of Nose and Chin. Evaluating differences in eigenvalues for movement in height of nose and chin uses five different files: original egg, Nose-up, Nose-down, Chin-up and Chin-down. Figure 3.9 compares the eigenvalues of the original egg, Nose-up and Nose-down, and Figure 3.10 makes the same comparison between the egg, and Chin-up and Chin-down. These figures show that as the height of the nose or chin increases from the original position, the eigenvalues decrease in magnitude and vice versa.

3.4.3.4 Changing Position of Nose and Chin. Five files similar to those used in height changes of the nose and chin are required for this analysis: original egg, Nose-xpos, Nose-xneg, Chin-xpos, and Chin-xneg. Figure 3.11 compares the energy matrix eigenvalues of egg, Nose-xpos, and Nose-xneg, and Figure 3.12 shows a similar comparison between egg, Chin-xpos, and Chin-xneg. No noticeable trend in eigenvalues appear. This can be explained by evaluating the variogram, Equation 3.10, to see how the variables change in this instance.

The sill of the variogram for these eggs remains constant because the z coordinates do not change from egg to egg, but the h's do change. The variogram function depends upon both the values of the sill, and h. Note that if h is greater than the range, a, the variogram is constant, the value of the sill. There are only a small number of h's which are less than the range, therefore only a small number of $\gamma(h)$ values change in the $\mathbf{L}$ matrix (refer to Equations 3.72 and 3.73) and thus in $\mathbf{L}_n^{-1}$. Therefore, no noticeable change in the bending energies can be detected between these eggs.

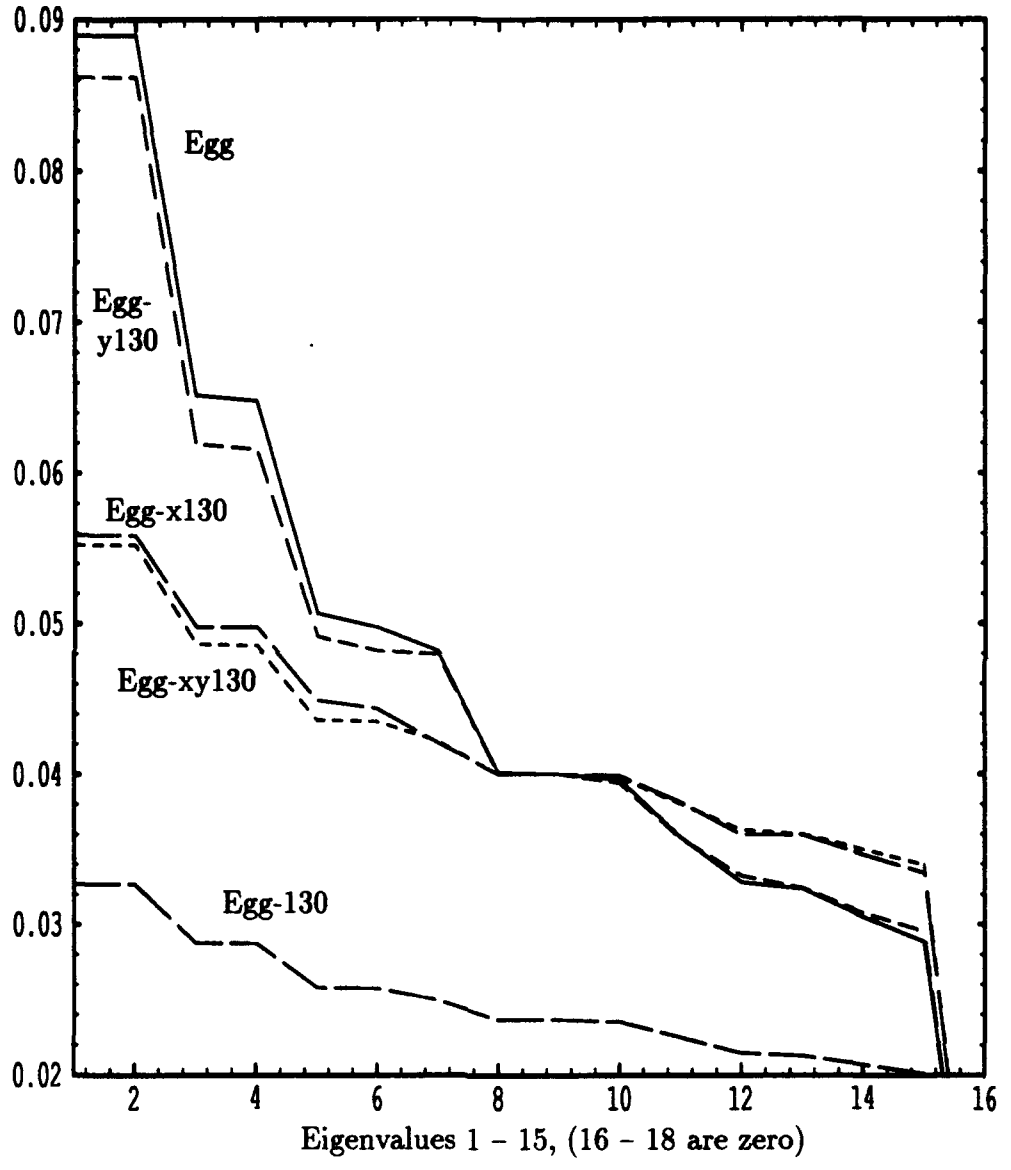


Figure 3.8 Egg Scaling Effects Trend: Trend appearing when eigenvalues of the energy matrix from scaled egg data are compared. The vertical axis represents the magnitude of the eigenvalues.

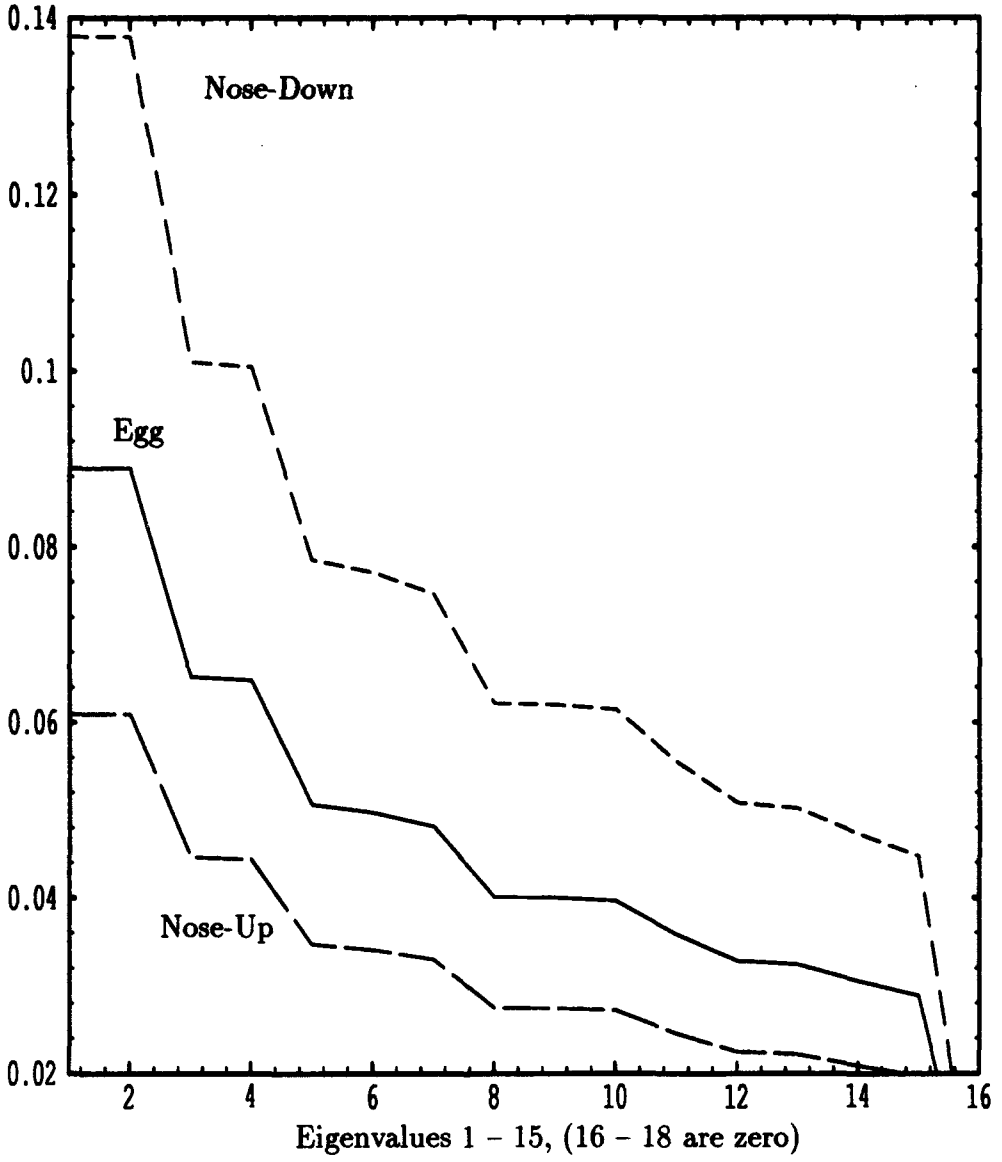


Figure 3.9 Egg Nose Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

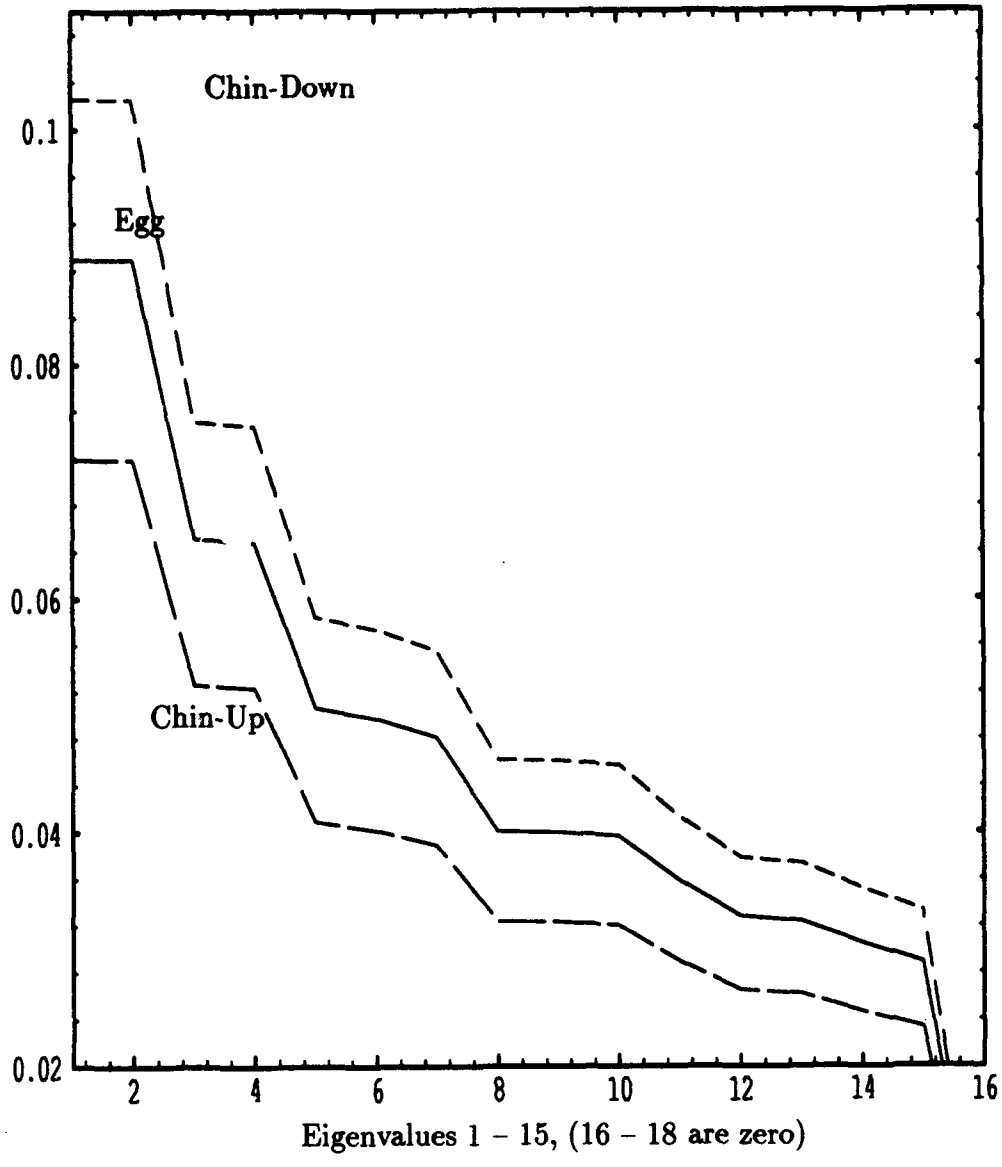


Figure 3.10 Egg Chin Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

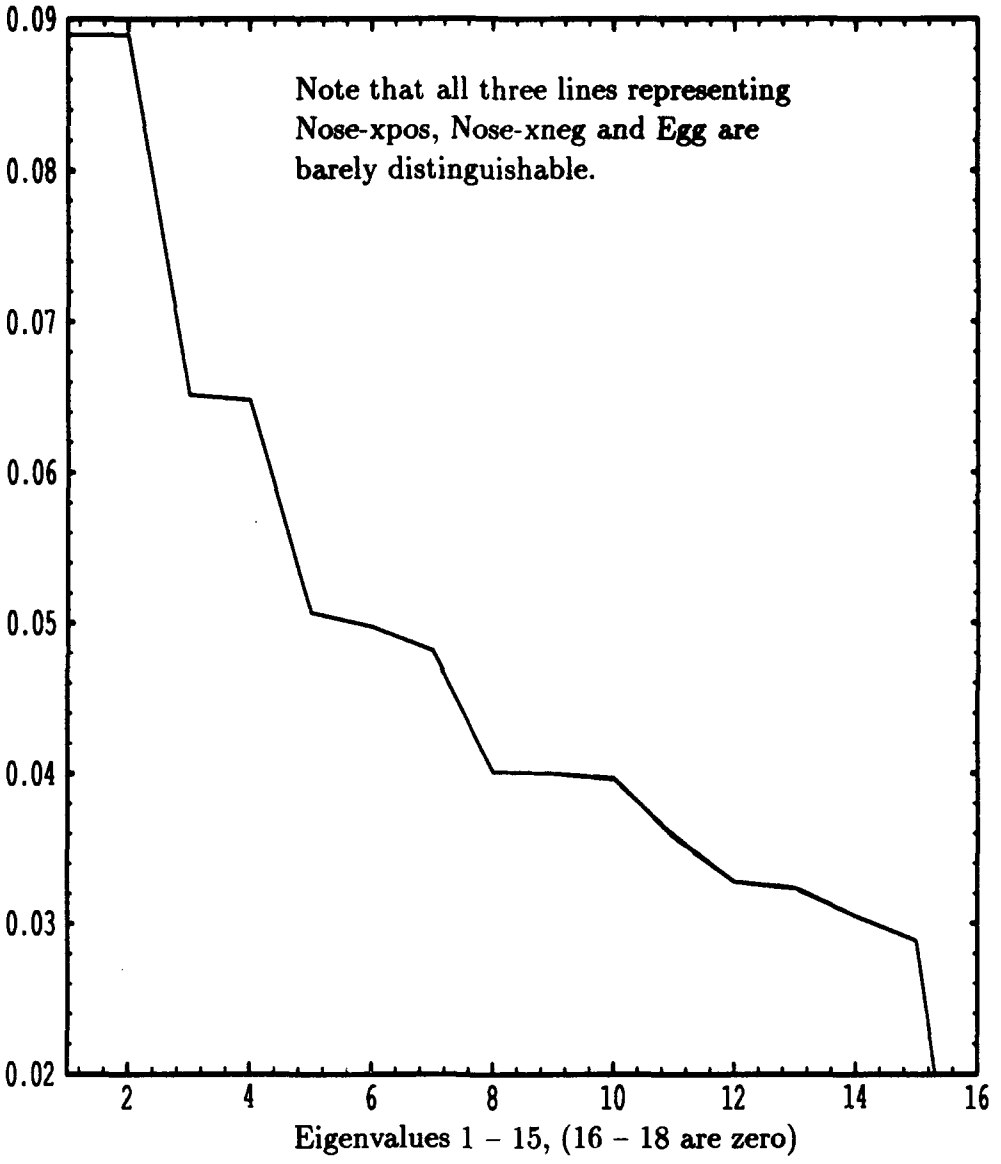


Figure 3.11 Egg Nose Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

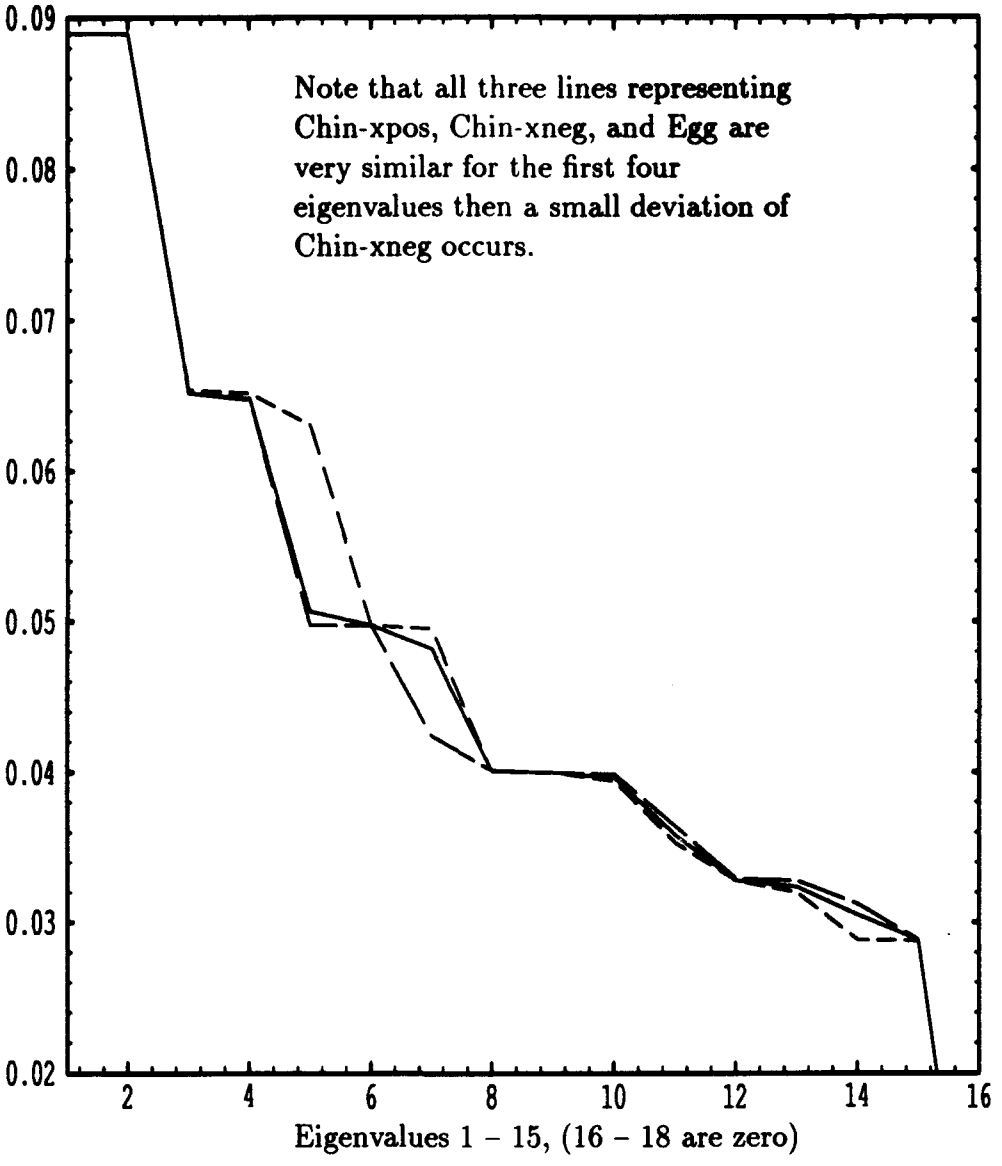


Figure 3.12 Egg Chin Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

3.4.3.5 Trend Validation. The trends depicted by modification in the form of a simple figure are validated on many other modifications to the egg and even to seal areas (Appendix B). Some of these modifications include adding more points to the circumference of the ellipse, where eggs having 20, 22, 26, and 30 points are created. Other modifications include: adding two more points within the ellipse to represent the corners of a mouth (Megg), and changing the heights of points along the ellipse to simulate 'depth' (Degg). Seal areas used in the validation are original seal area data digitized from plaster casts. The same changes in form itemized in Section 3.4.3 are made to all the new modifications and seal areas, and the same trends appear. Appendix B also contains eigenvalue plots of these modifications similar to plots shown in this section.

The comparisons made in this section demonstrate that the eigenvalues of the energy matrix contain information which characterizes the seal area of a face. These differences can be determined in an experiment such as those conducted in this section, but a problem occurs when trying to compare two seal areas. The eigenvalue variances between two seal areas may be due to several differences: one seal area being smaller, the nose/chin being lower, or maybe a combination of the two. These differences have the same effect on the eigenvalues from the energy matrix, $\mathbf{L}_n^{-1}$. The eigenvalues of the energy matrix do contain information which characterizes the seal area, but this may not be enough of all the information contained in $\mathbf{L}_n^{-1}$.

3.4.4 Energy Matrix Summary. The energy matrix has the ability to characterize the form of a face using spacial correlation. Each energy matrix is unique to a face, which identifies it as a source for features selection and comparison. The main advantage to using the energy matrix is that faces can contain an unequal number of data points. The disadvantage is the existence of a large data set, combined with the uncertainty of how to identify the critical features for comparison.

One method of data reduction is spectral decomposing the energy matrix into its eigensystem, which is unique to each matrix. The verification by reconstruction of the faces using bending energy modes and the validation methods using the egg, both justify using 'all' the eigenvalues of the bending energy matrices. However, clustering necessitates that the number of features used for comparison be equal, and dimensionality constraints require that the number of features be some minimum to justify using a greater number of clusters. The next chapter describes clustering techniques, fuzzy clustering in particular, and addresses these issues.

3.5 Form Characterization Summary

The energy matrix completely and uniquely describes the form of each face by kriging the seal area data. Kriging accounts for the anisotropy and the trend surface and then calculates the spacial correlation of the data.

Thin plate splines provide a technique for comparing figures but does not account for the spacial correlation of the data. The thin plate spline equations are the same form as the kriging equations except TPS uses the minimized energy function U and kriging uses the spacial correlation function γ. This research develops a true spacial correlation by combining kriging with the TPS methodology for comparing objects. The reference object is taken to be a flat plate and the energy matrix uniquely describes the bending of the flat plate to the object.

The energy matrix contains n^2 elements and must be reduced into features for clustering. The eigensystem decomposes the energy matrix into eigenvalues which, potentially, can be used as clustering features. Two techniques are discussed to determine the eigenvalues which best describe the form of a face. Bending modes combine each eigenvalue with its associated eigenvector in order to quantify the amount of bending associated with each eigenvalue. The Egg analysis investigates the effect on the eigenvalues of modifying a simplistic model of the face. Chapter IV

describes how features of the energy matrix are clustered to form groups of similar faces.

IV. Clustering

Clustering is the grouping of objects into characteristic classes. For example, a person without any prior knowledge of military aircraft may characterize them upon first sight as being either large or small, or either single- or multi-engined. It is often useful categorize sets of objects into groups with similar characteristics. Clustering algorithms have been developed to objectively accomplish this task. The two broad categories of clustering are fuzzy and crisp. Crisp clustering refers to 'classical' clustering where an object can belong to one and only one cluster. Fuzzy clustering utilizes fuzzy logic to account for situations where the same object may belong to more than one cluster. Both types of clustering are discussed in the balance of this chapter.

4.1 Crisp Clustering

Crisp clustering algorithms sort data into mutually exclusive and collectively exhaustive groups based upon common features. Implementing a crisp clustering procedure involves several considerations, including

- the number of clusters
- the number and type of features
- how to decide if a sample is in a cluster
- the user's definition of good clusters.

Each of the preceding terms deserves some further explanation.

4.1.1 Number of Clusters. The number of clusters depends on the available data and the application. Returning to the military aircraft example, *a priori* knowledge that the aircraft are either fighters or bombers yields the information that the data set is composed of only two clusters. If prior knowledge such as this is not

available, then it is desirable to construct an algorithm which searches out the optimal number of clusters. A common method of identifying this optimal number uses an iterative scheme to evaluate different numbers of clusters. After each iteration, the algorithm provides a calculated value of a 'criterion of optimality.' When the algorithm has terminated, the criteria can be compared to decide upon an optimal number of clusters. The criteria are commonly referred to as validity measures (2).

4.1.2 Feature Selection. The features of a data set are the characteristics used in a clustering algorithm to decide how the clusters are formed. The number and types of features, like the number of clusters, are characteristics which are not controlled by a series of cut and dry equations. Feature selection is very much an art, requiring experience, intuition, and common sense. Features should be chosen such that data samples are readily distinguishable from each other. In the design of the MBU-5/P oxygen mask, facial length and width from the 1950 anthropometric study were used as features. In the design of the MBU-12/P, facial length alone from the 1967 anthropometric study was the feature used for clustering. Features can also be derived from physical characteristics using transformations such as Fourier analysis or eigensystem analysis. The coefficients from these methods are used as features in a new, transformed feature space. Returning once again to the military aircraft example, it is desirable to locate the distinctive features which most easily separate 'fighters' from 'bombers.' Possible choices include wingspan, take-off weight, and maximum velocity. Any of the these features would probably provide more accurate clusters than using a feature such as the age of the pilots. It becomes apparent from this simple example that the type of features selected directly impacts the performance of the clustering algorithm. The choice of the type and number of features defines the feature space where the clusters exist. Choosing multiple features creates multiple dimensions in the 'feature space.' It also becomes apparent that the larger the number of distinguishing features which are available for use in the clustering, the more accurately the clusters are defined (54).

4.1.3 Decision Function Selection. A core element of any crisp clustering routine is deciding in which cluster an object belongs. The algorithm accomplishes this task by utilizing decision functions, also known as discriminant functions. Decision functions can be similarity measures which utilize a distance function. If sample A is closer to cluster 1 than cluster 2 in some multidimensional feature space, then the decision is made that A belongs to cluster 1. Some commonly used measures of similarity are Euclidean distance, Mahalanobis distance, and Tanimoto measure (54). The choice of a decision function is dependent upon the type of data samples and the desired results from the clustering algorithm. If there are known distributions within the data which can be exploited, the Mahalanobis distance, which incorporates the data covariance matrix, may be appropriate. If there are no known distributions, Euclidean distance may be appropriate. If the data is in binary format, the Tanimoto measure may yield the best results. Like feature selection, decision function selection is dictated heavily by the nature of the data and the experience of the programmer.

4.1.4 The Curse of Dimensionality. The allowable number of features, number of clusters and number of data samples used in a clustering algorithm are closely related. A trade-off exists which is not easily defined in mathematical terms.

> Unfortunately, it has frequently been observed in practice that, beyond a certain point, the inclusion of additional features leads to worse rather than better performance....The basic source of the problem can always be traced to the fact that the number of design samples is finite. However, analysis of the problem is both difficult and subtle. Simple cases do not exhibit the experimentally observed phenomena, and more realistic cases are difficult to analyze (21).

Many analyses have been conducted to derive some concrete rules relating dimensionality and sample size. The results of these studies provide general rules which are often contingent upon assumptions on the number of clusters, the probability distribution of the data, and the desired classification error rate for each particular

clustering algorithm. Raudys and Jain have recommended a sample size exceeding 1.2 times the dimensionality of the feature space for a two class problem (44). Cover has suggested the number of samples needs to be at least twice the dimensionality of the feature space for a two class problem (15). Tou and Gonzalez recommend at least ten times the number of samples per feature dimension (54). The relationship between number of features, number of clusters, and sample size can be grossly stated as:

$$\text{Number Clusters} \times \text{Number Features} \times \text{SF} \leq \text{Number Objects} \tag{4.1}$$

The 'SF' is termed the safety factor and refers to the multiplier needed to achieve reliable clusters. Per the above references, the SF has a usual range of between 1.2 to 10. More features make it easier to discriminate between classes, but the classifier needs enough data samples to support the additional features. Proof of these relationships abound (32, 44) but usually entail restrictive assumptions on the number of clusters, the distribution of the data, and the type of classifier. When designing any clustering or classification algorithm, it is important to keep these considerations in mind.

4.1.5 Performance Measures. Given a data set large enough for the dimensionality of the feature space and an acceptable maximum number of clusters, an optimal number of clusters within this maximum can be determined. As with decision functions, there exist many different measures of optimality. A commonly used performance index is the sum of squared errors (54) given by:

$$J = \sum_{j=1}^{N_c} \sum_{x \in S_j} \| x - m_j \|^2 \tag{4.2}$$

where:

x = data samples,

N_c = number of clusters,

S_j = set of samples belonging to the j^{th} cluster, and

m_j = mean of all values of the j^{th} cluster.

The minimization of this index gives tightly packed clusters. Other performance measures include cluster separation distances, cluster hypervolumes, and average squared distances between samples in different cluster domains (2). This is far from an exhaustive list. Many different performance measures have been developed for different applications. Again, the performance index chosen is dependent upon the data sample and the goals the clustering algorithm is attempting to achieve. According to Tou and Gonzalez, "..the performance of a particular cluster-seeking algorithm is dictated largely by the cleverness of its designers in extracting valuable information from the data being analyzed" (54).

If the samples from the data set are not easily separated into clusters, fuzzy logic can be incorporated into the clustering algorithm. The following sections give a brief overview of fuzzy logic, how it can be used in clustering, and a comparison between fuzzy clustering and crisp clustering.

4.2 Fuzzy Logic

Fuzzy logic is a system devised by Lotfi Zadeh in 1965 to quantify the uncertainty in a system (63). It is used in many applications including decision making, controls, and pattern recognition. Fuzzy set theory is the application of fuzzy logic to traditional set theory. It differs from classical (also known as crisp) sets in one key area. Whereas in crisp sets, an object is only allowed to be in one set, in fuzzy sets, an object is allowed to belong to more than one set. The degree to which an object partially belongs to a set is called its membership in that set and is usually depicted by μ. The definition Zimmerman gives is: "If X is a collection of objects denoted gener-

ically by x then a fuzzy set A in X is a set of ordered pairs: $A = \{(x,\mu_A(x))|x \in X\}$ where $\mu_A(x)$ is called the membership function of x in A" (64). The memberships of fuzzy sets quantify the ambiguity inherent in the real world. Rarely do real world objects have a single function or identity. An FB-111 Aardvark is both a fighter and a bomber and a KC-10 Extender is both a refueler and a cargo aircraft. Paul Newman is an actor, entrepreneur and a race car driver. Crisp sets are a subset of fuzzy sets since all memberships in a crisp set are either zero or one. Mathematically, crisp membership functions are defined as $\mu_A(x) \in \{0, 1\}$. Consequently, the membership function for fuzzy sets represents a degree of association with a particular set and is defined as $\mu_A(x) \in [0, 1]$.

The memberships $\mu_A(x)$ (where $x \in X$) of the points x in the set A can take different forms, four of which are graphically shown in Figure 4.1. The first representation is a discrete pairing of points within the set and the point's corresponding membership. If A is defined as {numbers close to 10}, then a possible fuzzy description of this set is A = $\{(x,\mu_A(x))|x \in X\}$ = {(7,0.1),(8,0.5),(9,0.8),(10,1.0), (11,0.8),(12,0.5),(13,0.1)}. The further a number is from 10, the smaller the value of its membership. Numbers not listed in this fuzzy set are assumed to have a membership of zero. Graphically, this fuzzy membership function is shown in Figure 4.1a. The set A = {numbers close to 10} can also be represented by a continuous function, such as $A = \{(x, \mu_A(x))|\mu_A(x) = (1 + (x - 10)^2)^{-1}\}$ (see Figure 4.1b). Two other common continuous membership functions are the trapezoidal (Figure 4.1c) and the triangular (Figure 4.1d).

4.2.1 Membership vs. Probability. It is important to note that memberships are not probabilities. Although memberships are restricted to be between zero and 1.0, there is no restriction on the area under the curve being 1.0 as in probabilities. For example, while a KC-10 Extender is both a cargo aircraft and a refueler, based on the type of mission or time spent accomplishing each mission it can have a membership of 0.8 refueler and 0.4 cargo aircraft. Another major differ-

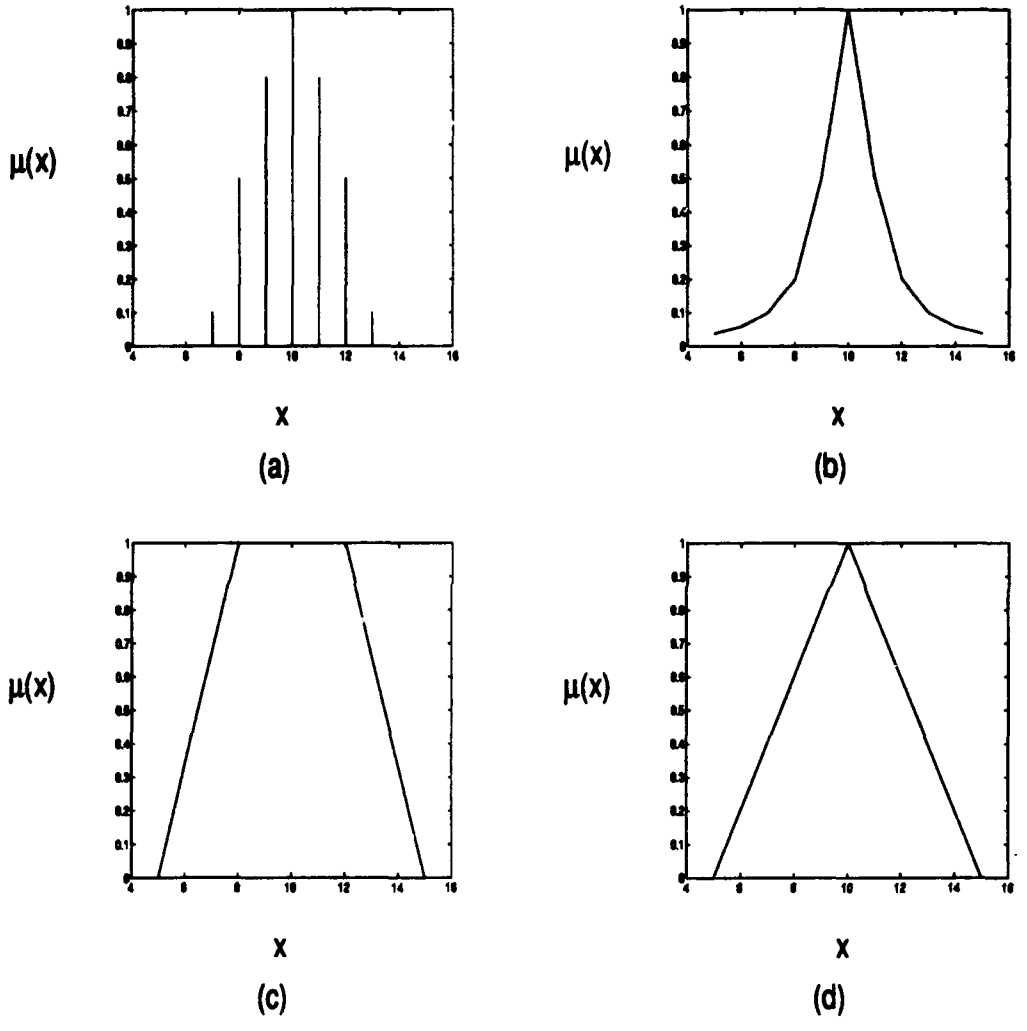


Figure 4.1 Fuzzy Membership Functions (a)Discrete membership (b)Continuous membership (c)Trapezoidal membership (d)Triangular membership.

ence between probabilities and memberships is that probabilities measure ambiguity in binary, 'either-or' terms. A commonly used example which displays this is the thirsty drinker. A man is lost in the desert and stumbles upon two bottles of liquid. Bottle A has a poison membership of 0.1. Bottle B has a probability of being poison of 0.1. Which bottle should he drink? If the membership function for poison represents the percentage of poison in the liquid, the membership of bottle A tells the man the liquid is 10% poison while the probability of bottle B tells the man there is a 10% chance the liquid is all poison. Assuming a 10 to 1 ratio of water to poison is nonlethal, the man should obviously choose bottle A. This example graphically illustrates the binarity of probabilities (2).

4.3 Fuzzy Clustering

Fuzzy clustering is the application of fuzzy set theory to clustering. The evolution of this application is a natural one. Crisp clustering is limited by the features the practitioner has available. The performance of the clustering algorithm largely depends on the resourcefulness of the practitioner to extract the most appropriate features from the data set. It is also up to the practitioner to develop a discriminant function that separates the samples into representative clusters. A challenge arises when, through no fault of the practitioner, representative features are not readily available. An equally challenging aspect of clustering is a data set which is not separable into definable clusters by a crisp discriminant function. Fuzzy sets allow the practitioner a margin of flexibility which makes clustering this data a viable exercise (64).

4.3.1 Cluster Membership. As in fuzzy sets, fuzzy clustering allows an object to belong to more than one cluster at a time. The degree to which an object belongs to a cluster is its membership in that cluster. Fuzzy clustering allows the practioner to decide if an object should belong to a cluster by the object's membership. Depending upon the application and the nature of the data, 'hard' membership

cutoffs can be decided upon to create crisp clusters. Objects with memberships less than these cutoffs can be considered indeterminant and not belong to any cluster.

4.3.2 Fuzzy Prototypes. The centroid of a particular cluster is sometimes referred to as the cluster's prototype. One advantage of fuzzy clustering is its ability to yield prototypes from data sets which by all appearances have none. Given a set of data which is not readily separable into n clusters, fuzzy clustering can identify which n entities best represent the distribution of the entire data set.

4.4 Fuzzy vs. Crisp Clustering

Fuzzy clustering and crisp clustering share many of the same qualities. They are both influenced by the type of data available and the curse of dimensionality. Also, they are both highly dependent upon the experience and knowledge of the practitioner in choosing features and discriminant functions. The one major difference between fuzzy and crisp clustering is the ability of fuzzy clustering to characterize the distribution of the data points in a more realistic manner than crisp clustering. Although it may suffice to say for many clustering applications that the objects are wholly in one cluster and not in another, it is an artificiality which sometimes can not be accepted.

4.4.1 Example: Fuzzy vs. Crisp Clustering. A classic example of fuzzy versus crisp clustering is the 'butterfly' (64). Figure 4.2a shows a set of 15 data points in two dimensional space. Figure 4.2b is how a crisp clustering algorithm might cluster these data points. A and B mark the centroids of each cluster. The data points in cluster **I** are completely in that cluster. In fuzzy terms, they have a membership of 1.0 in cluster **I** and a membership of 0.0 in cluster **II**. The memberships for this case are shown next to their respective points. Similarly, the points in cluster **II** have 0.0 membership in cluster **I** and a membership of 1.0 in cluster **II**.

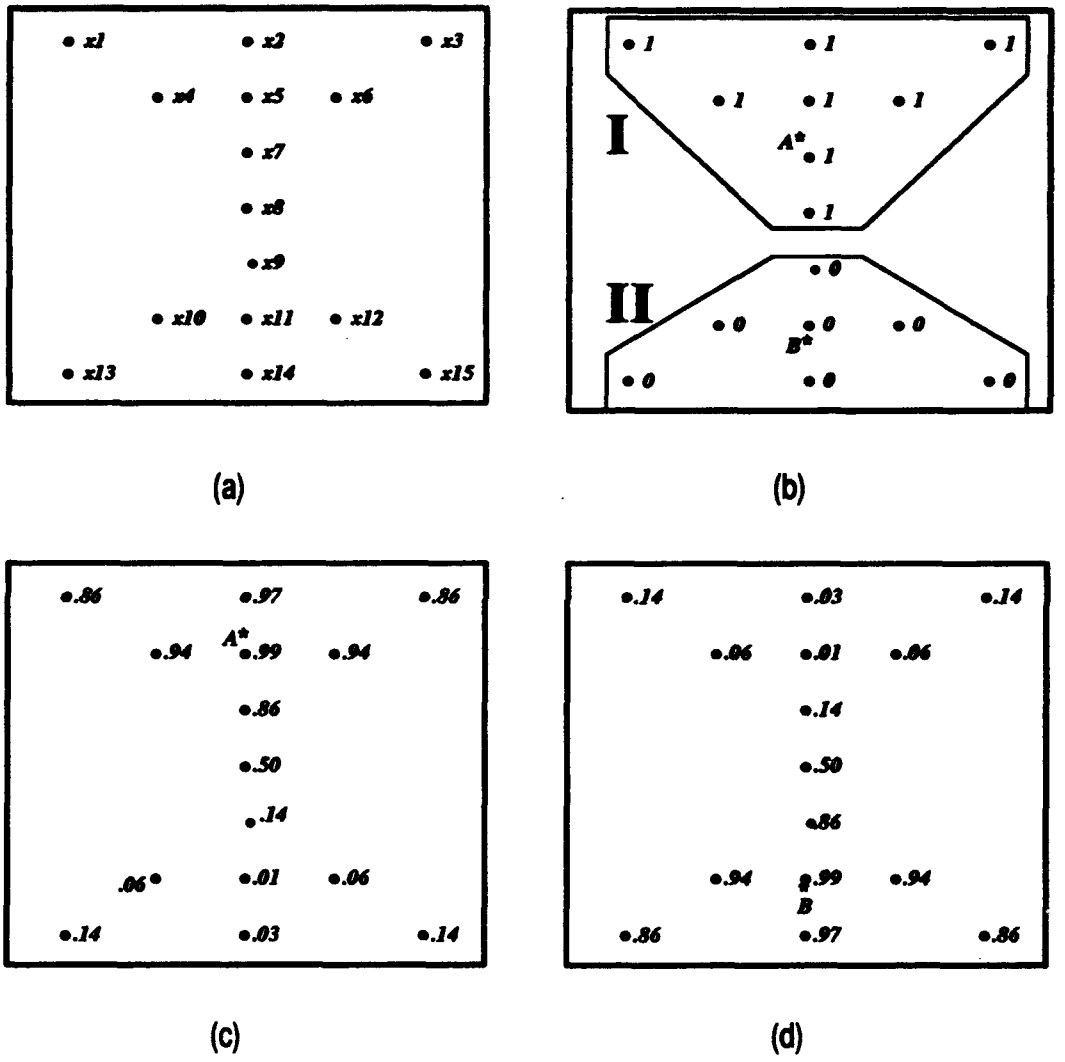


Figure 4.2 Butterfly Data Points (a)Labeled data points (b)Crisp clustering with equivalent fuzzy memberships for cluster **I** (c)Fuzzy clustering with respect to centroid **A** (d)Fuzzy clustering with respect to centroid **B**.

The two clusters from this crisp clustering routine may truly represent how these data points would be distinguished in actual practice. More likely, however, the pattern of their distribution does not indicate such a precise division into two sets. Figures 4.2c and 4.2d show how a fuzzy clustering algorithm might evaluate the same set of data. The values next to each data point are the memberships associated with a particular cluster. Figure 4.2c is the membership values for points compared to cluster center A and Figure 4.2d is the membership values for cluster center B. Assuming the features chosen truly represent the samples, the fuzzy clustering yields a more accurate and flexible characterization of the data. If the practitioner decided to create two clusters from the fuzzy clustering, he could use an 0.80 membership cutoff to decide which clusters data points belong to. Using this threshold, the clustering would come out the same as in the crisp clustering with the exception of point x_8. Point x_8 has a 0.50 membership with respect to centroid A and a 0.50 membership with respect to centroid B. These memberships indicate point x_8 is half in the top cluster and half in the bottom cluster. Using a 0.80 cutoff, the point x_8 would be considered indeterminant. The flexibility indicated by this simple example could be extended to more complicated situations.

4.4.2 Fuzzy Clustering Summary. The attributes of fuzzy clustering allow a measure of flexibility which is applicable to this research. Fuzzy clustering is useful for this research because small differences in faces are not easily distinguishable. It has the unique ability to assign memberships to objects of different clusters and determine the theoretical centroid of a cluster. By identifying the memberships of objects it becomes an easy task to identify the object with the highest membership in each cluster. This is the object which is closest to the theoretical centroid and best represents the cluster.

4.5 Fuzzy Clustering Algorithm

This section describes the fuzzy clustering algorithm used in this research. There are many crisp clustering algorithms which have been altered to use fuzzy logic. The algorithm chosen for this research is the fuzzy c-means algorithm (62). Fuzzy c-means provides a relatively straight-forward method for analyzing data of an unknown distribution.

4.5.1 Fuzzy C-Means. The fuzzy c-means algorithm derives its name from the fact that the user must input the number of clusters, c. It assigns memberships to objects based upon the objects' Euclidean distances to the fuzzy centroids of the clusters. In addition to the number of clusters, the user must input the feature vectors, $\mathbf{x}_j$, representing each of the sample objects.

The first step in the fuzzy c-means algorithm is to determine the number of clusters. The methodology of choosing the number of clusters is explained in Section 4.6. The membership of face j in cluster i is initialized for each face and cluster. The initial memberships are uniform random numbers ranging from 0 to 1. The memberships are then normalized such that:

$$\sum_{i=1}^{c} \mu_{ij} = 1 \tag{4.3}$$

where c is the number of clusters.

The features describing face j are put in the vector $\mathbf{x}_j$. The centroid $\mathbf{v}_i$ of each cluster is calculated by:

$$\mathbf{v}_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m \mathbf{x}_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \tag{4.4}$$

where $m > 1$ is the fuzziness index and n is the number of faces. As an exponential weight, m influences the membership function. For very large values of m the

membership of every face in every cluster will be $1/n$. This is referred to as a fuzzy membership function. Conversely, small values for m lead to higher memberships and a more crisp clustering. When using random data, there is no real 'concentration' around the centers of the clusters, so a value of $m = 2$ is usually used. The squared Euclidean distance of each face $\mathbf{x}_j$ from a cluster centroid $\mathbf{v}_i$ is:

$$d^2(\mathbf{x}_j, \mathbf{v}_i) = (\mathbf{x}_j - \mathbf{v}_i)^{\mathrm{T}} \mathbf{A} (\mathbf{x}_j - \mathbf{v}_i) \tag{4.5}$$

where $\mathbf{A}$ is a symmetric, positive-definite weighting matrix, usually the covariance of the data set. The matrix $\mathbf{A}$ is used to determine the shape of the cluster calculated by the fuzzy c-means algorithm. Since the feature space represents an isotropic bending energy matrix, $\mathbf{A}$ is the identity matrix and the shape of the clusters are equally sized hyper-spheres (64).

The membership functions are then updated using:

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(\mathbf{x}_j, \mathbf{v}_i)}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{c} \left(\frac{1}{d^2(\mathbf{x}_j, \mathbf{v}_i)}\right)^{\frac{1}{m-1}}} \tag{4.6}$$

The updated membership functions will be normalized as shown in (Eq. 4.3).

The objective function to be minimized for fuzzy c-means is:

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m (\mathbf{x}_j - \mathbf{v}_i)^{\mathrm{T}} \mathbf{A} (\mathbf{x}_j - \mathbf{v}_i) \tag{4.7}$$

The objective function measures the distance weighted by the memberships. Equations 4.4 and 4.6 are iterated until the objective function J_m converges to a local minimum. The result of the optimization is high memberships at small d^2 and vice versa. According to Xie and Beni, the fuzzy c-means algorithm always converges to a strict local minimum (62). The global minimum is found using an exhaustive search technique by initializing μ_{ij} a large number of times.

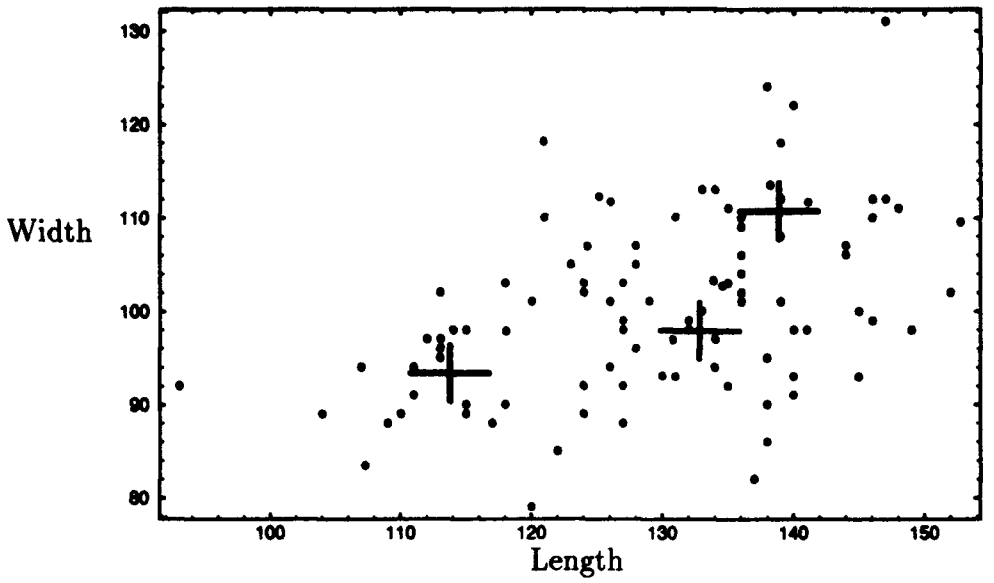


Figure 4.3 Length and width of 99 faces and resulting cluster centroids.

The final clusters produced by the fuzzy c-means algorithm are more stable then crisp clustering algorithms, even if the data appears to be a random scatter. The results of the clustering are c clusters centered at $\mathbf{v}_i$. High values of membership represent faces that are strongly related to the cluster and vice versa.

4.5.2 Example. The c-means clustering method is applied to the length and width of each face to explain the procedure. Clustering on linear facial data does not produce well defined clusters (24). However, it is useful to look at length and width clustering because two feature (i.e., two-dimensional) clustering can be easily visualized.

The data set consists of 99 faces. Figure 4.3 has 99 dots representing the length and width of each face. The faces are clustered based on 3 clusters and the resulting centroids are shown as crosses in Figure 4.3.

Figures 4.4, 4.5, and 4.6 give the memberships of each face in each of the three clusters. The memberships are indicated by the size of the dots. Points close to the cluster centroids have large memberships (bigger dots), points close to another

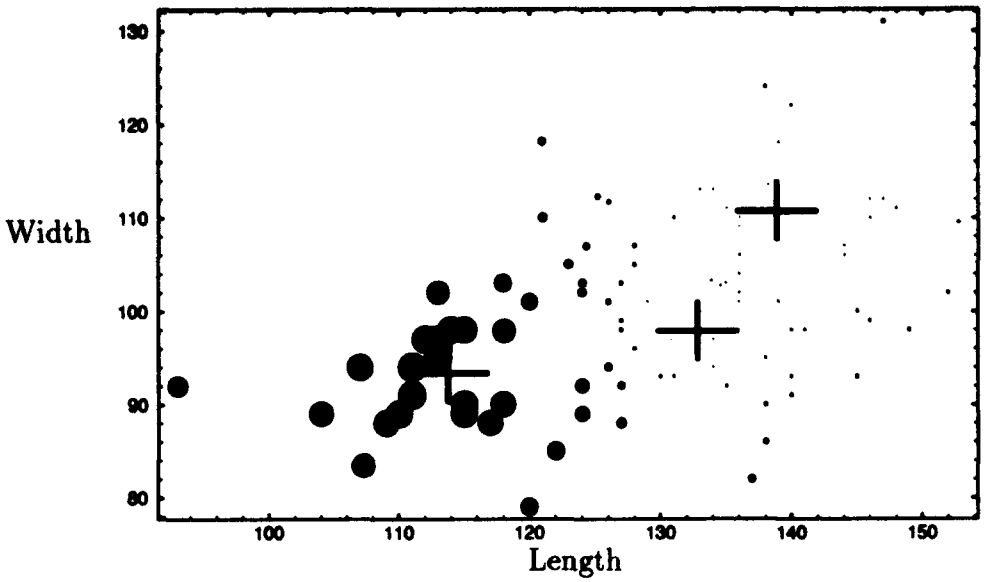


Figure 4.4 Membership of 99 faces in lower left cluster based on length and width clustering.

cluster centroid have low memberships (smaller dots), and points equally distant from the three centroids have equal membership in each of the three clusters.

4.6 Number of Clusters and Features

The first step in executing the fuzzy c-means clustering algorithm is to define the number of clusters. While mathematical operations exist to define an optimum based upon a specified criterion, none of these are applicable to this research. The number of clusters and features is arrived at from an iterative process based upon dimensionality constraints for the limited sample size of 99 faces. Section 4.1.4 discussed the relationship between the number of features, the number of clusters, and the size of the data sample. If the number of features is increased, the algorithm can more accurately classify the data samples. With increasing the number of clusters, the entire data sample will be better represented by the increase in the number of cluster prototypes. This research experimented with many different combinations of clusters and features in an effort to find a compromise which would yield accurate clusters which cover an adequate size of the sample data.

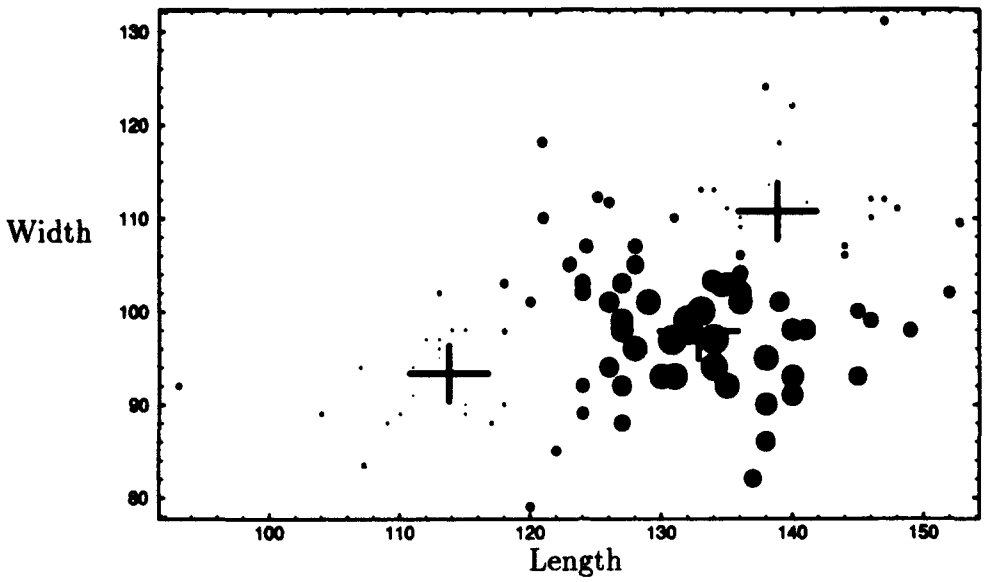


Figure 4.5 Membership of 99 faces in center cluster based on length and width clustering.

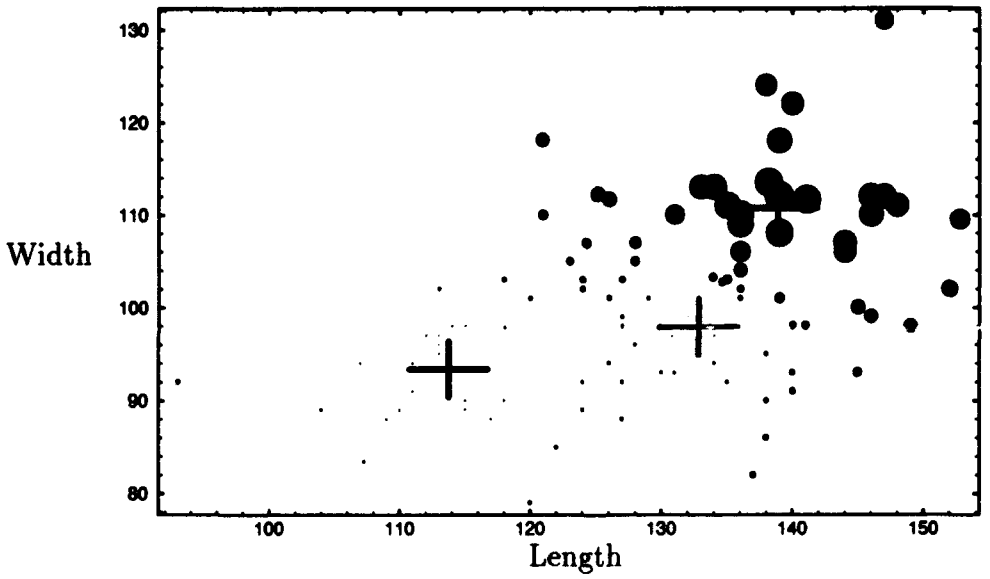


Figure 4.6 Membership of 99 faces in upper right cluster based on length and width clustering.

In experimenting with the available data sample of 99 faces, this research is targeted for the production of six prototype masks using five features. This combination of number of clusters and number of features yields a conservative safety factor of 3.30. Clustering using the fuzzy c-means algorithm is done on two through six clusters in an effort to gain some insight into how the data samples behave for different values of c.

4.7 Facial Feature Selection

One of the most challenging aspects of any clustering algorithm is feature selection. Feature selection entails choosing which charateristics of the data best separate the data into the desired clusters. This is a twofold challenge when the desired clusters are not known *a priori*. Such is the case with this research. Features which have been used in past research are landmark dependent and have not yielded satisfactory results. Distances between landmarks on the human face are highly uncorrelated and difficult to consistently define. The use of the energy methods discussed in Chapter III provides an alternative to reliance on landmarks. Recall that the form of an object is described by the energy matrix, $\mathbf{L}_n^{-1}$. This matrix is unique for each object and provides a new set of possible features; the eigenvalues of the $\mathbf{L}_n^{-1}$ matrix are also unique and can be used as a possible set of features as well. As discussed in the previous section, this research has been limited to clustering with five features. However, both $\mathbf{L}_n^{-1}$ and its eigenvalues have substantially more than five elements. The challenge becomes one of feature extraction and reduction. The balance of this section describes the alternative ways to extract five features for clustering from these potential feature sets. These reductions are broadly divided into free form and fixed form methods. Free form methods extract features directly from the eigenvalues of $\mathbf{L}_n^{-1}$. In contrast, fixed form methods extract features from $\mathbf{L}_n^{-1}$ matrices that have been reduced and transformed into a new feature space.

4.7.1 Free Form. The bending energy matrix is created without the use of landmarks. Therefore, each face contains a different quantity of data depending upon its relative size. That is, the total number of data points, n, is not constant from face to face because small faces have less data than large faces, $n_{small} < n_{large}$. This results in different bending energy matrix sizes. The free form method of feature space determination identifies the unique features of each individual face from the uneven data sets. This section describes an application of the free form method which uses the eigensystem of each bending energy matrix to create the feature space for clustering.

Each seal area can be described by its bending energy matrix. The bending energy matrix is very large, consisting of n^2 features. Dimensionality constraints dictate reducing this information to an amount compatible with the clustering routine. The key is to retain the vital form characteristics of the seal area during this reduction. As was shown in Section 3.4, the eigensystem of the bending energy matrix, $\mathbf{L}_n^{-1}$, uniquely describes the form of a face (5). The eigenvalues represent the magnitude of the bending energy, and the related eigenvectors identify the location of the bending. This concept was validated with the bending modes in. One interpretation of the bending modes is that large eigenvalues could be indicative of high bending energies, which occur in areas of high curvature, such as around the nose. These large eigenvalues could also be related to more localized differences in the shape of a face compared to a flat plane. Conversely, small eigenvalues may indicate larger scale differences or global changes, like the size of a face. The bending modes also verify the reconstruction of a face from the eigensystem using the spectral theorem. In choosing which five of these eigenvalues to use as a feature set, a decision is made by the research team concerning the relative importance of global form versus local variations. A large variety of five dimensional feature sets can be created from this data.

Global form characteristics are shown to be contained in the smaller eigenvalues of the $\mathbf{L}_n^{-1}$ matrix, whereas local variations are shown to be contained in the larger eigenvalues of the $\mathbf{L}_n^{-1}$ matrix. For this research, the objective is to choose features which separate the sample faces into groups of members who wear similarly sized oxygen masks. The challenge is to select the combination of global features and local features that best characterize a sample face with respect to mask fit. Both overall size and the localized area around the bridge of the nose are equally important. Unfortunately, the best combination of five eigenvalues is not easy to identify. Unless some other criterion for feature selection is created, trial and error appears to be the only method for identifying which five eigenvalues to use.

4.7.1.1 Free Form Summary. The advantage of using free form methods is that facial feature sets can contain an unequal number of features within each data sample. The disadvantage is the large number of possible features within each facial data sample and the lingering question of which are the critical features for comparison. The verification by reconstruction of the faces using bending energy modes justifies using 'all' the eigenvalues of the bending energy matrices. However, dimensionality constraints limit the number of features that can be used in the clustering routine. Clustering necessitates that the number of features used for comparison be equal. This does not mean the free form method is unworkable or has no value. The free form method is of significant value if a method could be developed to locate and extract five critical features. After much experimentation, a method to accomplish this was not found during this research. Another method to determine the feature space is reducing the raw data for each face to an equivalent quantity before calculating the bending energy matrix. This is the fixed form method. Using fixed form allows for the use of linear transformations to identify critical features for comparison.

4.7.2 Fixed Form. This section discusses the steps of feature reduction and extraction of the original data set. The first step is to reduce the data by sparsing the points on each seal line to a uniform number. Next, the data is transformed into a new feature space using a linear transformation. Uncorrelated features are then extracted from this transformed data to provide an alternative data set.

The experiments with the egg data and the bending modes show that there is indeed information in the energy matrix $\mathbf{L}_n^{-1}$ which is unique for each face form. It is also shown that some of this information is contained in the eigenvalues of the same energy matrix. By calculating the eigenvalues of $\mathbf{L}_n^{-1}$, a substantial reduction in the amount of data (from n^2 to n) is achieved. However, for a sample of only $n = 99$ objects, the curse of dimensionality dictates a feature space much smaller than even this reduced data set. The challenge is in deciding which elements from $\mathbf{L}_n^{-1}$ or which of its eigenvalues to use in the clustering algorithm.

4.7.2.1 Sparsed Data. In the original data collection, points are digitized every 2 millimeters along the seal line of the plaster casts. A first effort at reducing the dimensionality of the data set is to delete redundant points along this curve without changing the curve's original shape. Using the sparsing function on the Pro/ENGINEER CAD software (see Section 5.1.2.3), two thirds of the original smoothed data points can be deleted without affecting the shape of the original seal curve. To aid in the future comparison between faces, the seal curves are sparsed into a uniform number of points.

The number of points chosen to represent each seal curve is 39. Empirical testing show that most seal curves start to change shape when reduced to between 25 to 30 points. 39 points is chosen because it allows a substantial reduction in the amount of data and a margin of confidence in maintaining the original shape of the seal curve. With the addition of the 14 spacial delineators, the size of $\mathbf{L}_n^{-1}$ is 53×53, yielding 2809 components and 50 non-zero eigenvalues. A further discussion

of sparsing is found in Section 5.1.2.3. At this point, the same eigenvalues which emphasize global form and local variations can be extracted from this sparsed data set as in the free form data set.

Although this represents a significant reduction in the dimensionality of the data set, the sparsed $\mathbf{L}_n^{-1}$ matrix still contains too much information for clustering with a limited sample size of 99 faces. The hypothesis for this sparsed set of data remains the same as for the original, free-form set of data. The sparsed $\mathbf{L}_n^{-1}$ or its eigenvalues can be used to represent the form of the sample face. If it is known which elements of $\mathbf{L}_n^{-1}$ or which eigenvalues of $\mathbf{L}_n^{-1}$ 'best' characterize the face for clustering, a further data reduction and possible clustering could be accomplished. A means of achieving this is the Karhounen-Loeve Transform (KLT).

4.7.2.2 KLT. The KLT is a linear transformation which creates an orthogonal, uncorrelated data set. The KLT has seen much use in the field of image processing (26). It is referred to by other names such as principal component analysis and the Hotelling Transform. For the purposes of this research, the real strength of using the KLT is in the ordering of the transformed data set. The order of the coefficients of the transformed data set represents an ordering of features in decreasing variance. In other words, the first transformed feature has the greatest variance, the second transformed feature has the second greatest variance, and so on. In using the KLT for feature selection, it is assumed that a 'good' feature dimension is defined as one having much variance. The more variance along the dimensions of the feature space, the easier it will be to decide upon possible clusters. Figures 4.7 and 4.8 graphically display this concept. Features A and B were used to cluster the data set in Figure 4.7. Features C and D were used to cluster the same set of samples in Figure 4.8. The larger variances of features C and D afforded a clearer choice of choosing the prototypes, indicated by the X's.

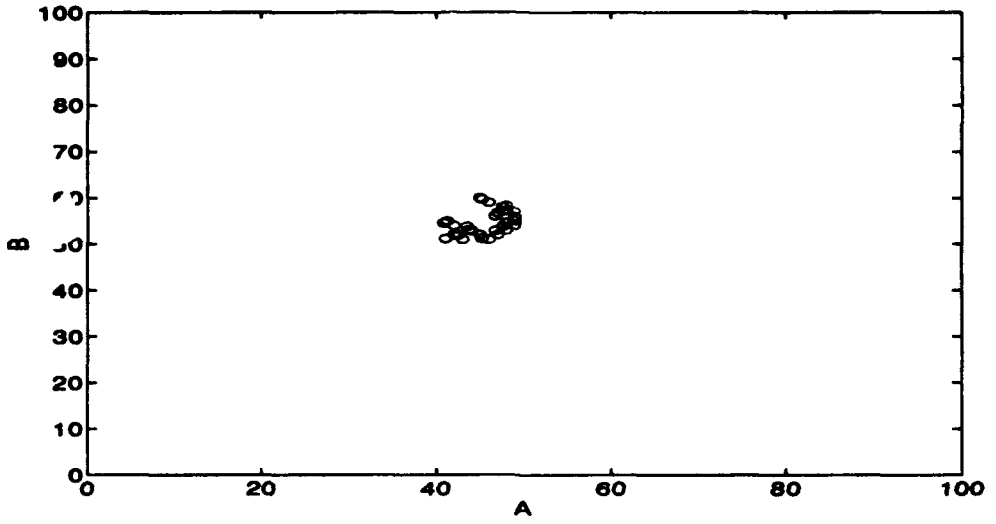


Figure 4.7 Small variances within features.

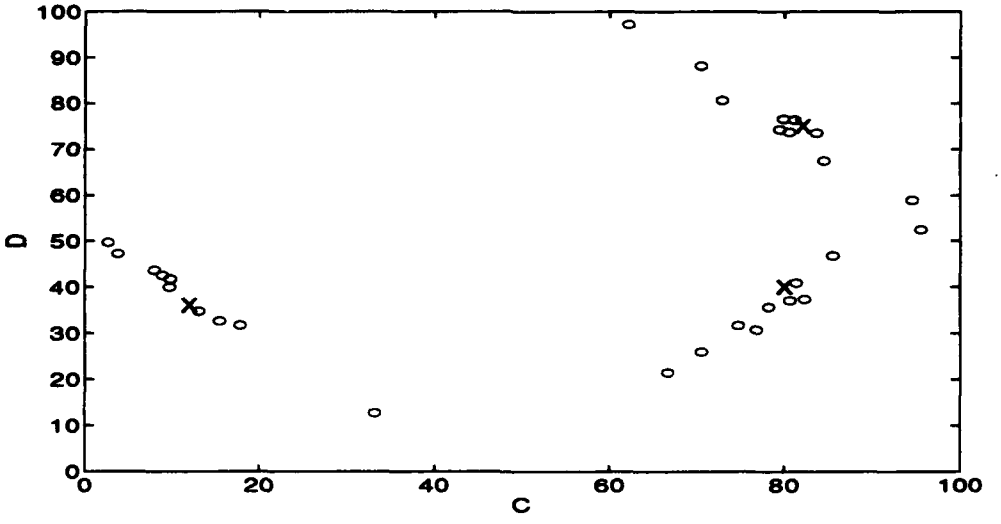


Figure 4.8 Large variances within features.

The steps in the KLT are as follows:

Given:

m data samples

n points in each sample

$\mathbf{\Phi}_i$, where $i = 1, ..., m$ mean-zero feature vectors

Step 1. Let the matrix $\mathbf{A}$ be composed of all the feature vectors $\mathbf{\Phi}_i$:

$$\mathbf{A} = [\mathbf{\Phi}_1 | \mathbf{\Phi}_2 | \ldots | \mathbf{\Phi}_m] \tag{4.8}$$

Step 2. Since the $\mathbf{\Phi}_i$'s are mean-zero, define the covariance matrix ($\mathbf{C}$) of $\mathbf{A}$ to be:

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathrm{T}} \tag{4.9}$$

Step 3. Calculate the eigenvectors ($\mathbf{x}_i$) and eigenvalues (λ_i) of $\mathbf{C}$.

Step 4. Create a matrix of eigenvectors corresponding to the ordered eigenvalues (from highest to lowest):

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{bmatrix} \tag{4.10}$$

Step 5. Calculate the KLT coefficients by multiplying the ordered eigenvector matrix by the original data vectors:

$$\mathbf{X}\mathbf{\Phi}_i = \mathbf{\Omega}_i \tag{4.11}$$

where $\mathbf{\Omega}_i$ is the vector of the KLT coefficients of the original feature vector $\mathbf{\Phi}_i$. The $\mathbf{\Omega}_i$'s are ordered in such a way that the first component over all the $\mathbf{\Omega}_i$'s has the greatest variance, the second component has the second greatest variance, and so

on. The KLT coefficients are the transformed data set for use as a feature set in a clustering algorithm.

The KLT is applied to two distinct sets of data: $\mathbf{L}_n^{-1}$, and its eigenvalues. Using the KLT on the eigenvalues of $\mathbf{L}_n^{-1}$ is a direct application of the aforementioned process. However, use of the KLT on the entire $\mathbf{L}_n^{-1}$ requires further explanation. To create the feature vectors $\mathbf{\Phi}_i$, each $\mathbf{L}_n^{-1}$ is 'vectorized.' That is, each matrix is converted into a vector by appending the second row of the matrix onto the first row, the third row onto the end of the first/second combination, and so on. If the original matrix is $n \times n$, the vectorized version is n^2 long. The $\mathbf{A}$ matrix composed of all these $\mathbf{\Phi}_i$'s is $n^2 \times m$, where m is the number of different $\mathbf{L}_n^{-1}$ matrices. Since the covariance matrix is created via $\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$, $\mathbf{C}$ has dimensions of $n^2 \times n^2$. In the case of the sparsed data set, this is a matrix of size 2809×2809. Finding the eigenvectors and eigenvalues of such a matrix is not a trivial task. There is a linear algebra substitution which can be used to reduce this matrix into a more reasonable size:

Step 1. Define $\mathbf{L} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$ (size $m \times m$)

Step 2. The eigenvectors ($\mathbf{e}_i$) and eigenvalues(μ_i) of $\mathbf{L}$ are defined by

$$\mathbf{L}\mathbf{e}_i = \mu_i \mathbf{e}_i. \tag{4.12}$$

Step 3. Substituting $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{L}$,

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{e}_i = \mu_i \mathbf{e}_i. \tag{4.13}$$

Step 4. Multiplying each side by $\mathbf{A}$,

$$\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{e}_i = \mu_i \mathbf{A}\mathbf{e}_i. \tag{4.14}$$

Step 5. Substituting $\mathbf{C} = \mathbf{A}\mathbf{A}^T$,

$$\mathbf{C}(\mathbf{A}\mathbf{e}_i) = \mu_i(\mathbf{A}\mathbf{e}_i). \tag{4.15}$$

Therefore, $\mathbf{C}$ has eigenvectors $(\mathbf{A}\mathbf{e}_i)$ and eigenvalues μ_i.

For this reduction to have any meaning, the eigenvalues and eigenvectors of the total covariance matrix $(\lambda_i,\mathbf{x}_i)$ would have to be equivalent to the eigenvalues and eigenvectors of this new substitution $(\mu_i,\mathbf{A}\mathbf{v}_i)$. It can be shown that if two matrices are similar, they have the same eigenvalues (50). The definition of similarity is: Two matrices $(\mathbf{C},\mathbf{L})$ are similar if $\mathbf{L} = \mathbf{A}^{-1}\mathbf{C}\mathbf{A}$. For the case at hand, the proof of similarity is:

$$\begin{aligned} \mathbf{L} &= \mathbf{A}^T\mathbf{A} \\ \mathbf{A}\mathbf{L} &= \mathbf{A}\mathbf{A}^T\mathbf{A} \\ \mathbf{A}\mathbf{L} &= \mathbf{C}\mathbf{A} \\ \mathbf{L} &= \mathbf{A}^{-1}\mathbf{C}\mathbf{A} \end{aligned}$$

Therefore, $\mathbf{L}$ and $\mathbf{C}$ are similar. They have the same non-zero eigenvalues, and their eigenvectors are related by $\mathbf{v} = \mathbf{A}^{-1}\mathbf{x}$ (50).

4.7.2.3 Fixed Form Summary. The KLT can be applied to both the eigenvalues of $\mathbf{L}_n^{-1}$ and the entire $\mathbf{L}_n^{-1}$ matrix. The KLT can also be applied to a selected number of eigenvalues from the unsparsed, free form data set. This yields sets of ordered, orthogonal, uncorrelated features. The question of 'which' features to choose is now answered. One suggestion is to choose the top n_f features from each data set, where n_f is dependenent upon the size of the sample and the number of clusters.

It is important to note at this point that the only reason the KLT is able to be utilized is that the data sets have all been reduced to the same number of

points. It would have been impossible to create the **A** matrix if all the Φ_i's had been different sizes. The acknowledgement of this point brings the logic train back full circle. If there exists a method for extracting the 'most characteristic' components or eigenvalues of the original, unsparsed $\mathbf{L}_n^{-1}$, the restriction on identical dimensionality between data samples would not apply.

4.7.3 Feature Selection Summary. The energy matrix $\mathbf{L}_n^{-1}$ uniquely characterizes the form of a face and provides an large number of possible feature sets. This section has presented a sampling of possible feature sets which could be used in a clustering algorithm. Finding which feature set yields the best results is dependent upon the application. It is an iterative operation in which selected possibilities are tested against each other and some criterion of optimality.

4.8 Clustering Summary

Clustering is a process which requires as much common sense as mathematical agility. There are many parameters that can be altered to affect the results of the clusters. Such considerations include the kind of clustering algorithm, the number of clusters, the type and number of features and the number of data samples. This research limits the fuzzy c-means clustering algorithm to no more than six clusters and the size of the data set is fixed at 99 faces. The number of features is chosen to be five as a conservative estimate of what the dimensionality constraints allow. This research concentrates its efforts on extracting a five-dimensional feature set from the $\mathbf{L}_n^{-1}$ energy matrix. Features are extracted from the eigenvalues of this matrix, the eigenvalues of the sparsed energy matrix, the KLT coefficients of sparsed energy matrix, and the KLT coefficients of the eigenvalues of the sparsed energy matrix. The following chapter addresses in more detail the selection of the feature set for clustering and how that feature set performs against the criteria established in this chapter.

V. *Application and Results*

This chapter discusses the methodology applied to develop better fitting aircrew oxygen masks (AOMs). Figure 5.1 shows the top-level process. The methodology begins by collecting the data and preparing it for form characterization. The form characterization process consists of analyzing the data structure, using universal kriging to obtain the bending energy matrix, $\mathbf{L}_n^{-1}$, and extracting features from the bending energy matrix. These features are used to group faces into similar sets, or clusters, with a c-means clustering algorithm. The face with the highest membership is selected from each cluster, and prototype masks are constructed to fit this set of high-membership faces. An evaluation is then made of how well the prototype masks fit other members of the cluster. The following sections explain each element of the process in greater detail. Throughout the chapter, subject 0002v is used as the sample data. Appendix C contains data outputs from each step in the data collection and preparation of face cast 0002v.

5.1 Data Collection and Preparation

This section explains the steps of the process used to collect and prepare the data (see Figure 5.2). It begins with collecting the sample data and continues through processing the three-dimensional coordinate data to prepare it for form characterization.

5.1.1 Data Collection. The plaster face casts used in fabricating custom MBU-5/P AOMs contain all the data needed for form characterization. The face casts were obtained from the Custom Mask Shop (CMS) of the 645th Medical Group, Aerospace Physiology Directorate, at Wright Patterson Air Force Base. As the only producer of custom-made, flight-certifiable masks in the Department of Defense, the CMS's fabrication process is unique. The next section reviews the custom mask

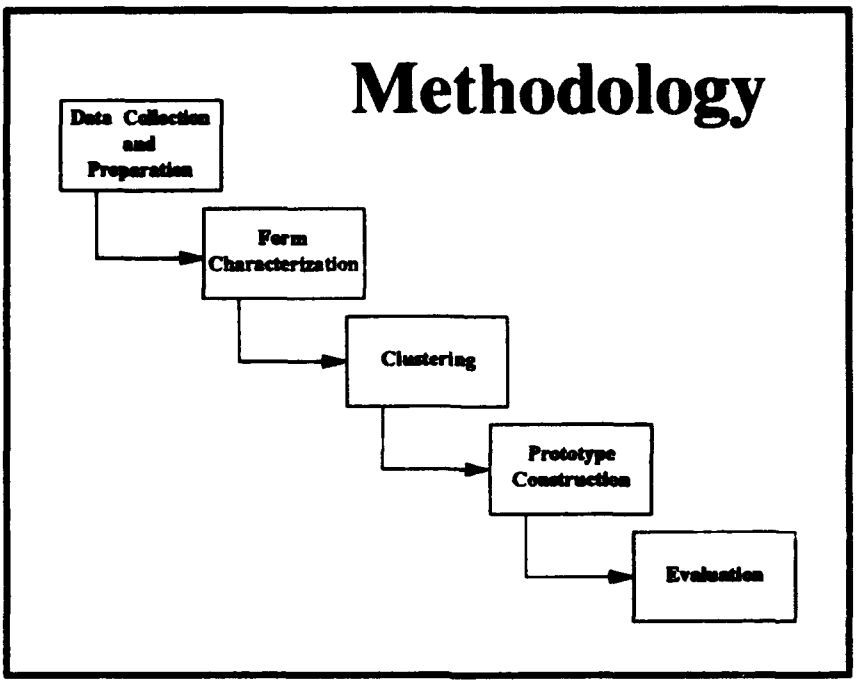


Figure 5.1 Methodology Process Flow Diagram.

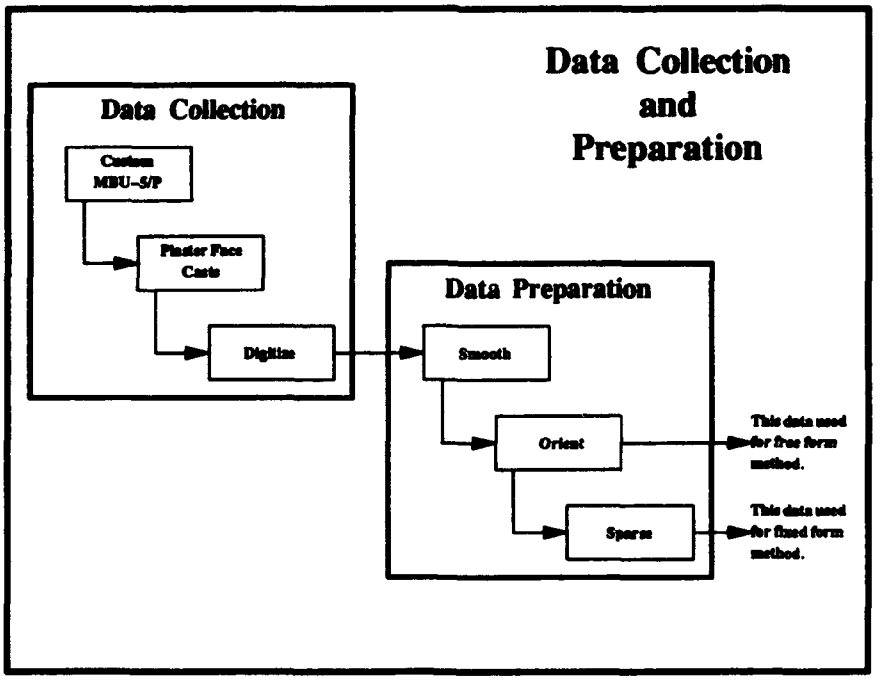


Figure 5.2 Data Collection and Preparation Process Flow Diagram.

Category I	Initial issue of custom mask
Category II	Latex faceform replacement
Category III	Latex faceform and hard-shell replacement
Category IV	Hard-shell replacement
Category V	Using an existing mask mold (plug) that conforms to the requesting individual's face cast

Table 5.1 Category Classifications for Custom MBU-5/P Masks.

making process to introduce the particular terminology and data associated with the process.

5.1.1.1 Custom Mask Process. The custom mask process begins with a request sent from any DoD organization, including the Army, the Navy, and the Marine Corps. The requests are classified into the five categories shown in Table 5.1 (57:¶ 3.b.5). Pictures of an entire standard MBU-5/P mask, latex faceform, hard-shell, plug, and plaster face cast can be seen in Figures 5.3 through 5.9.

In order to create a mask, the requesting organization must provide the CMS with a plaster face cast of the individual who, for some reason, can not be fit by one of the standard sized masks. No information arrives with the plaster face casts to indicate why a standard issue mask does not conform to the individual. The CMS personnel, who are experienced in fitting standard masks, surmise that most requests are made due to faces that are extremely small (mostly female), have large noses with very thick bridges, or are extremely flat.

The 'seal line', the line where the latex faceform contacts the face, is drawn according to Operating Instructions written by the CMS (60:¶ 3.b). Several different laboratory technicians drew the seal lines on the plaster face casts collected for this study, adding some variability into the form characterization process. The requirements for the finished latex faceform are: the mask will cover from the deepest part of the bridge of nose to 25.4 mm under the chin; 25.4 mm on both sides of the nostrils; and 31.8 mm on both sides of the mouth. Because the latex shrinks during

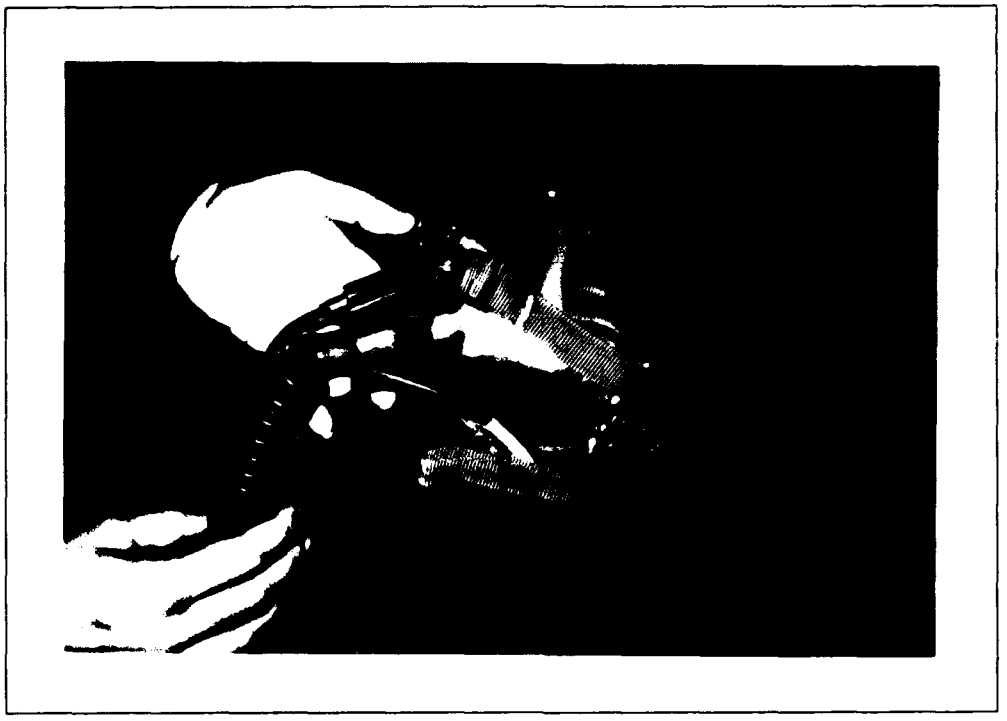


Figure 5.3 Standard MBU-5/P Mask.

Figure 5.4 Interior View of Standard MBU-5/P Mask.

Figure 5.5 MBU-5/P Custom Latex Faceform.

Figure 5.6 Unfinished MBU-5/P Custom Hard-Shell.

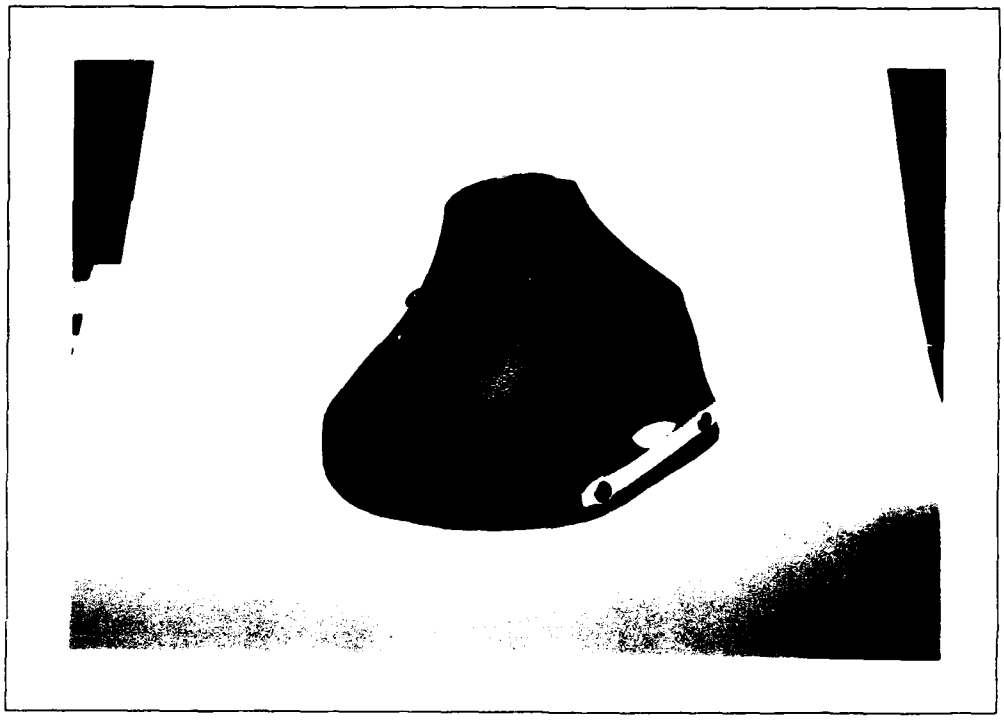


Figure 5.7 Finished MBU-5/P Custom Hard-Shell.

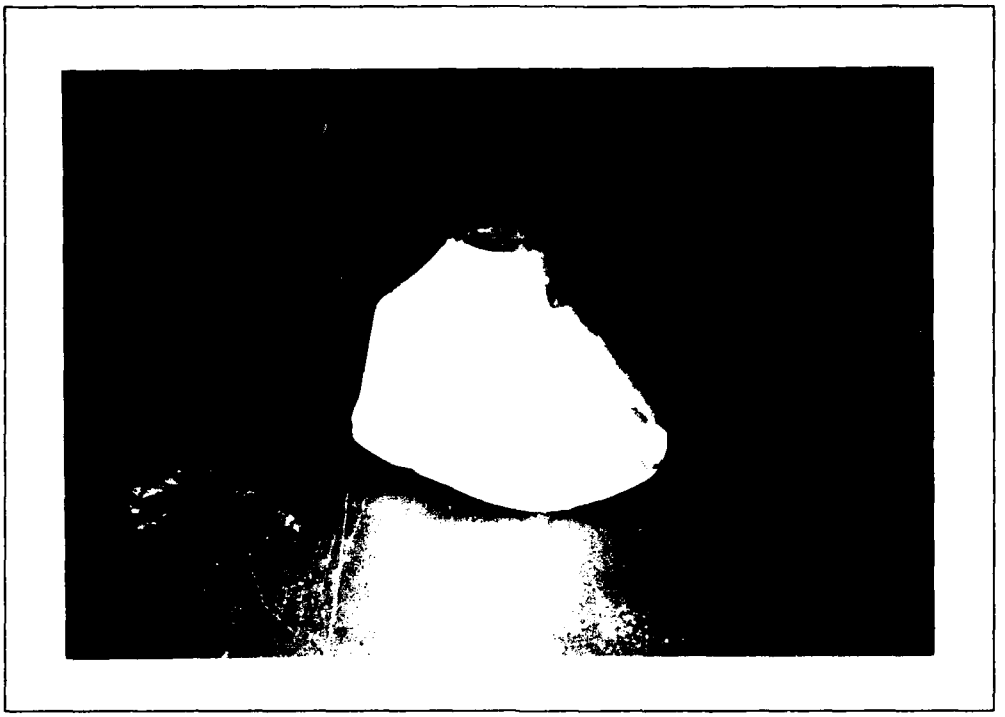


Figure 5.8 MBU-5/P Custom Plug.

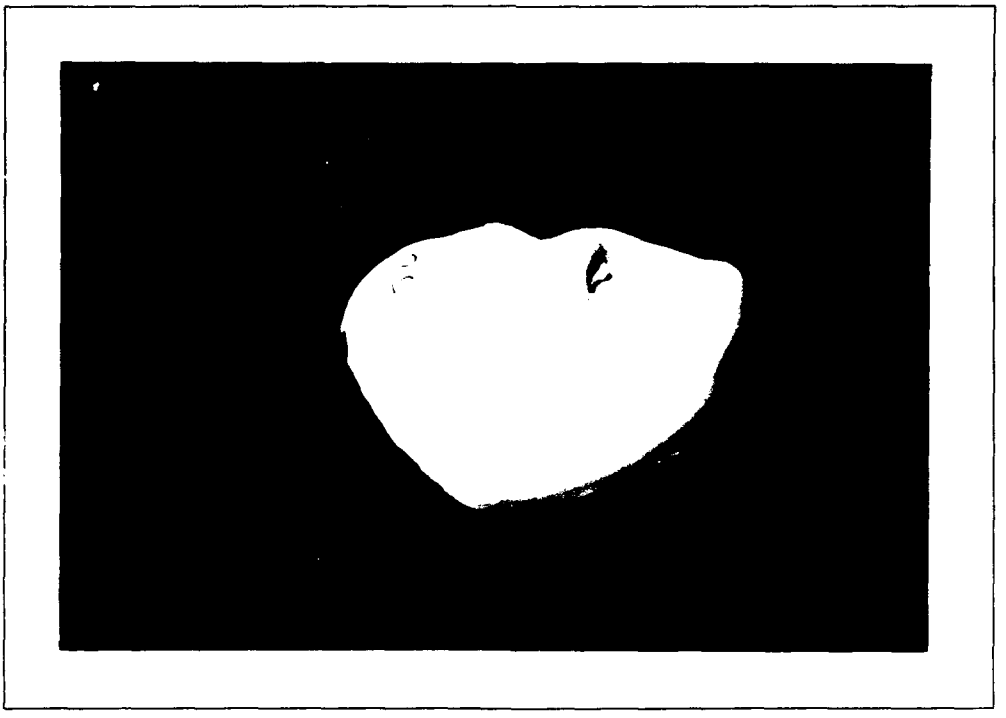


Figure 5.9 Plaster Face Cast: A plaster face cast used in MBU-5/P custom mask process.

the curing process, the mold or plug must be approximately six percent larger than the mask faceform, and the seal line is drawn accordingly. The ultimate goal is to make the plug as small as possible while still meeting the minimum finished mask size. Keeping the plug small eliminates as much dead space in the finished mask as possible, and so minimizes rebreathing of exhaled air trapped in the mask.

The plug creation is the most time consuming step of the custom mask process. First, clay is used to fill negative areas of the plaster cast within the seal line (the nostrils, mouth, and any indentations below the lower lip or on the chin), as depicted in Figure 5.10. The seal area is coated with petroleum jelly to reduce adhesion of the plug material to the plaster face cast. Figure 5.11 shows a partial hard-shell positioned over the plaster cast. The next step is to build a pocket of clay around the seal line until it reaches the bottom of the partial hardshell positioned over the plaster cast. The top of the hardshell, where the valve port is located, is the only opening. Hydrocal, also known as dental stone, is poured into this opening and allowed to harden. The hardened hydrocal, referred to as the plug, is removed from the plaster face cast. The hardshell and any residual clay are then removed from the plug. This plug is sanded and shaped appropriately with particular attention paid to areas which may block the sight of the user. Once shaped, two coats of paint (consisting of two components: polymide and epoxy) are applied to the plug to protect it (60). Figure 5.8 shows a painted plug. The plug is then dipped in an alcohol-based coagulant before dipping into the liquid latex. This dipping process is repeated several times to ensure a thick coat of latex covers the plug. During the latex curing, the plug must withstand a cold water leach of 21° C (70° F) for three hours, a 45 minute hot water leach at 60° C (140° F), and an oven temperature of 50° C (120° F) for four hours or 66° C (150° F) for three hours (58, 59).

Once the latex faceform has cured, it is removed from the plug and subjected to a four hour hot water leach at $50-60$° C ($120-140$° F) (59). Next, the appropriate sized hardshell is made and the entire mask is assembled (61). Plugs are kept by

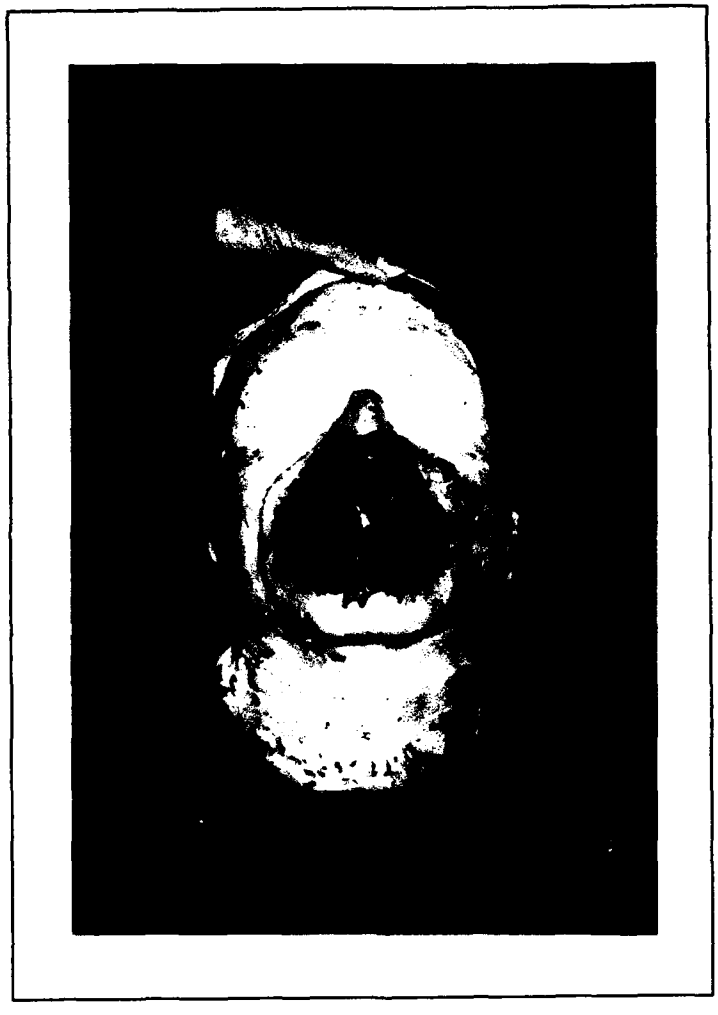


Figure 5.10 Plaster Face Cast With Clay: Negative areas of plaster face cast filled with clay during the manufacturing process of an MBU-5/P.

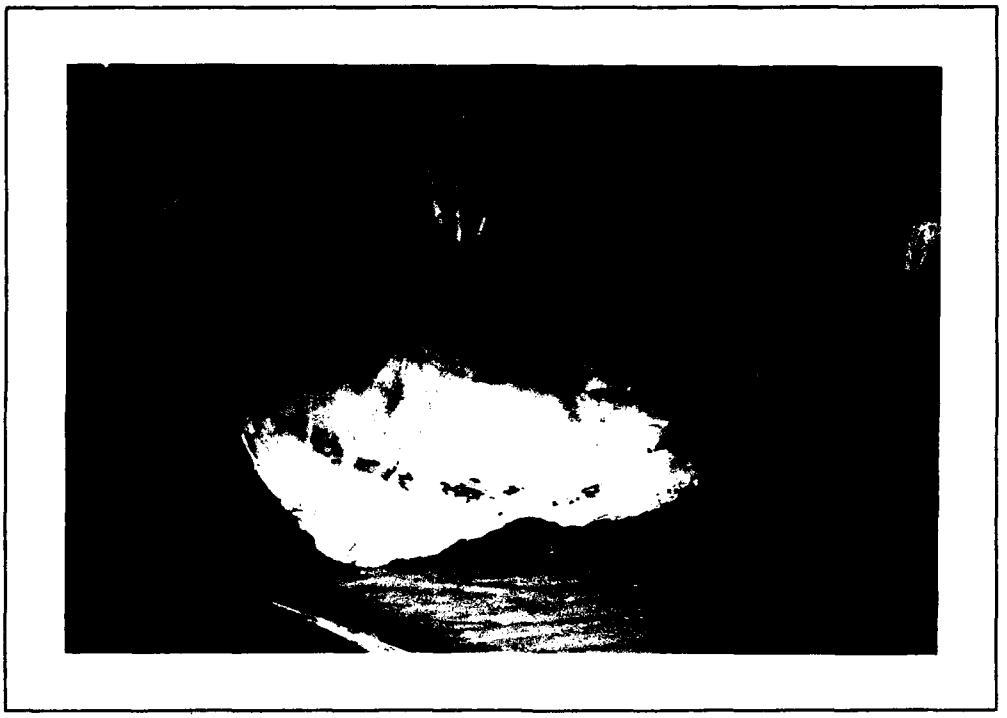


Figure 5.11 Positioning Hard-Shell: Positioning a partial hard-shell over a plaster face cast during the manufacturing process of an MBU-5/P.

the CMS for a minimum of six years. This is done for two reasons: first, there is the possibility of reuse (Category V); and secondly, the latex used in the manufacturing process loses its elasticity and becomes brittle after 18 months, requiring the faceform to be replaced. The plaster face casts are typically kept for only six months. There is not enough room to store all the casts received and six months was deemed an appropriate amount of time to ensure that a second mask can be made if the first one does not fit.

5.1.1.2 Plaster Face Casts. The plaster face casts stored at the CMS are the source of the data used in this research. After collecting the casts from the CMS, excess clay and hydrocal are removed so that the seal line and the area it encircles are clean. Once this cleaning is done, 14 'spacial delineators' are marked (see Figure 5.12). The delineators are chosen to help depict the shape of the face inside the seal area, especially in the hard to fit regions where it is vitally important that there be no air leakage past the mask-face seal. Eight of the 14 spacial delineators are located in the area of the nose, and the others are chosen to show the overall shape of the face, so that if a flat plate is deformed to pass through the points chosen, the shape of the face is adequately represented. The plaster face cast is now ready for digitization.

5.1.1.3 Digitize. After cleaning and marking the 14 spacial delineators (SD), the seal line and 14 SD are digitized using a METRECOM three-dimensional digitizer with 0.125 mm accuracy and version 1.01 of the Caliper 3-D software (23). The seal line is obtained by using 'stream mode', where an (x, y, z) coordinate is collected every two millimeters, and the SD are taken as points. For consistency, every face is digitized in the same manner: the seal line trace starts at the bridge of the nose and continues in a clockwise direction, and the SD are digitized in numerical order. The plaster face cast is not moved between the two digitizations; this ensures that the seal line and the 14 SD are collected in the same

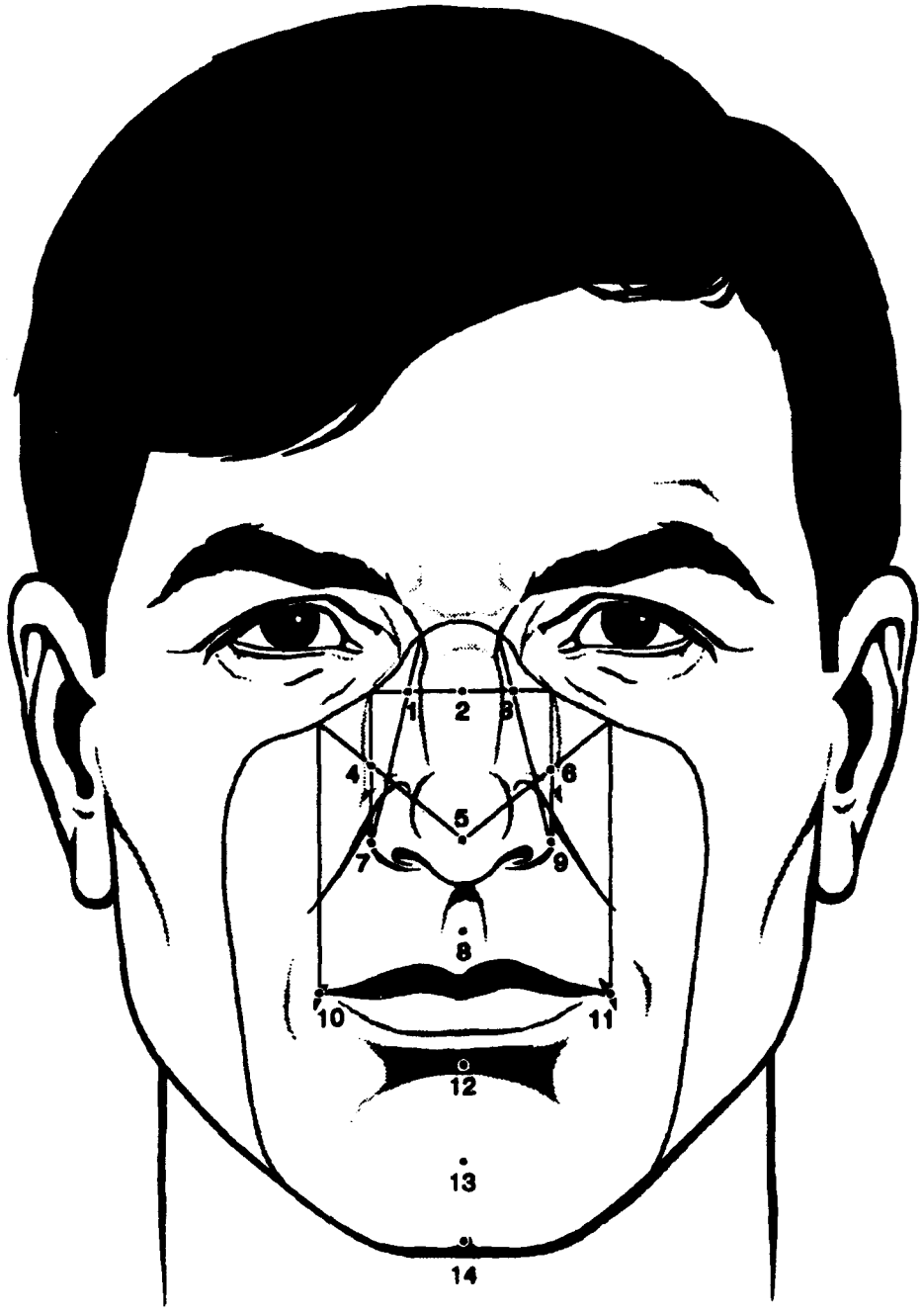


Figure 5.12 Spacial Delineator Sketch: Spacial delineators shown on a sketch of the face.

coordinate system. The seal line and SD are saved as two different files for computational convenience. These two files contain the raw data of the study and are ready to be processed in order to conduct form characterization analysis.

5.1.2 Data Preparation. The step-by-step process taken to create the data files needed to conduct form characterization analysis is outlined in this section. It describes the smoothing of the seal line, orientation of both data files, and sparsing of the seal line.

5.1.2.1 Smooth. As is mentioned in Section 5.1.1.3, hand digitization of the seal line is done in stream mode, where an (x, y, z) coordinate is taken every two millimeters. An entire file of seal line points consists of 130 - 180 points. A computer aided design program, Pro/ENGINEER (41), is used to highlight and remove digitization errors in the seal line files. When a seal line file is loaded into Pro/ENGINEER, the program uses a least-squares regression routine to fit a spline through the points and displays a closed loop showing the seal line. The curvature of the seal line file typically shows small inconsistencies, as seen in Figure 5.13. These inconsistencies occur due to the deformations such as bumps and holes in the plaster casts, and from operator error during digitization. Obviously erroneous points, such as those taken when the stylus pointer of the three-dimensional digitizer fell into a small hole on the plaster face cast, are removed. The next step is 'smoothing' the seal line by using a feature of the Pro/ENGINEER software which automatically averages the position of each point with those of its neighbors to either side. The points are relocated to these average positions, and a new curve passing through them is calculated and displayed. Comparing Figure 5.13 to Figure 5.14 illustrates how the irregularities in the original digitized seal line have been removed. The output from Pro/ENGINEER is a new data file of (x, y, z) coordinates defining the smoother curve (see Appendix C).

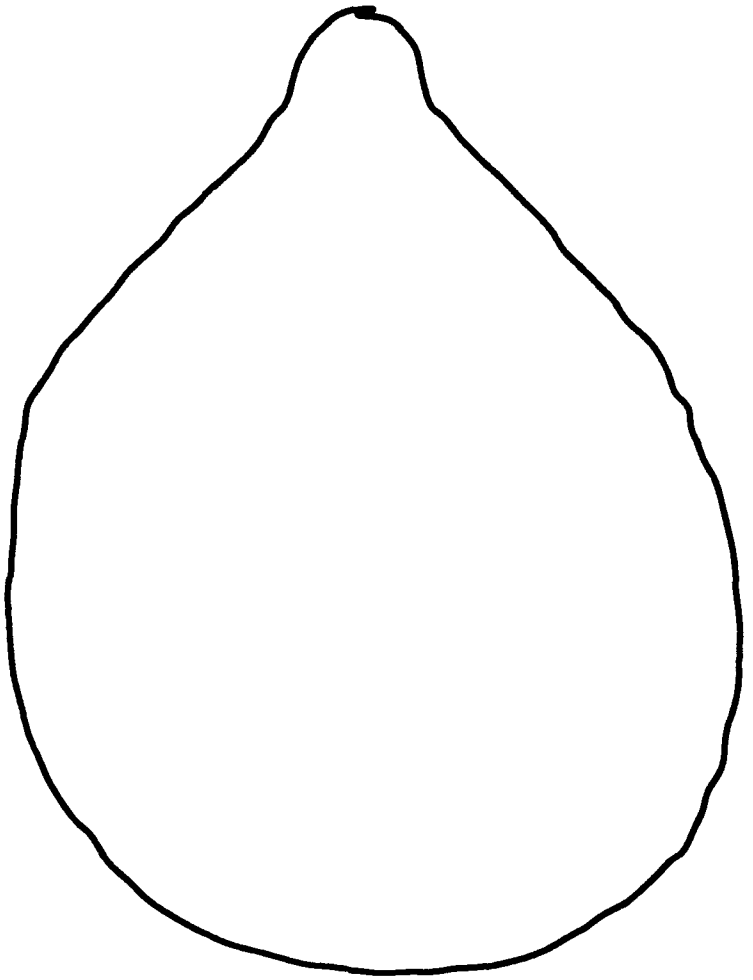


Figure 5.13 Unsmoothed Seal Line.

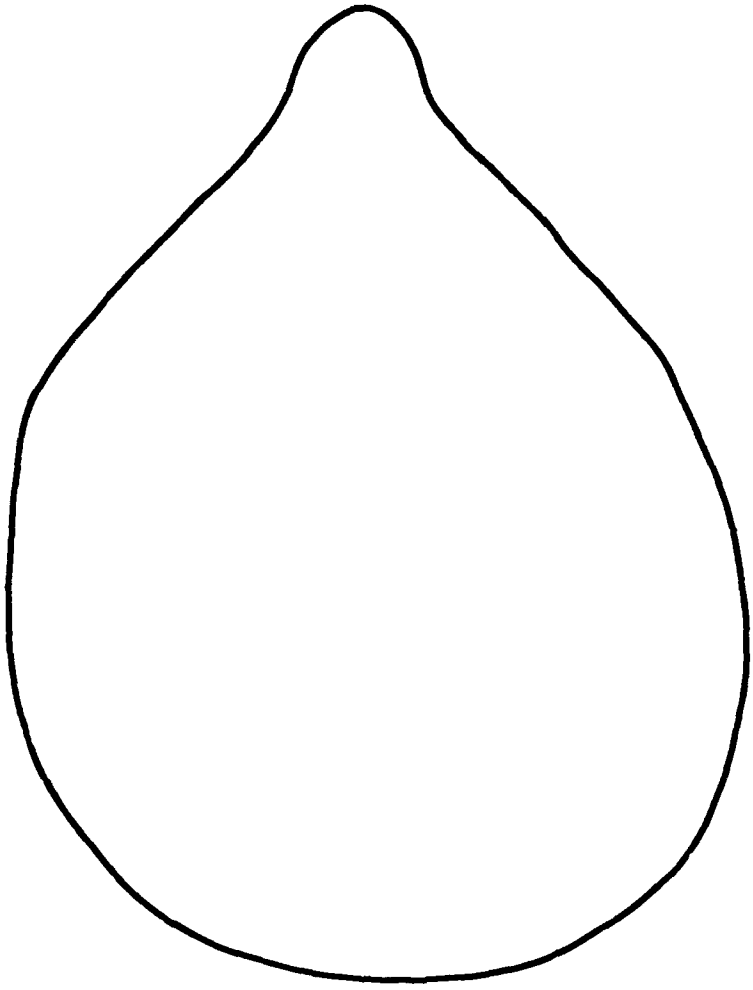


Figure 5.14 Smoothed Seal Line.

5.1.2.2 Orient. The form characterization method developed in Chapter III is independent of a common coordinate system, but for convenience in computations and visualization in CAD programs, a common orientation of the seal areas is needed. Historically the 'Frankfort Horizontal' plane is the standard plane of reference for craniometry using a human skull.

The Frankfort Horizontal plane is established by three landmarks on the human skull: the right and left *poria*, and the left *orbitale* (see Figures 5.15 and 5.16). The *porion* are the most lateral points on the roof of the external auditory meatus (bony ear-hole) and the *orbitale* is the lowest point on the lower margin of the orbit (eye socket) (34). This definition of the Frankfort Horizontal is modified for use on live subjects, because the *poria* are not distinguishable through the skin. A live subject's head is positioned so an imaginary straight line through the right *tragion* and right *infraorbitale* is horizontal to the ground, see Figure 5.17 (11). The plaster face casts do not contain any of the points needed to establish a Frankfort Horizontal plane, so another method, principal components analysis, is used to orient all seal areas in the same manner.

The smoothed seal line file of each face is joined with its SD file. This creates one file of (x, y, z) coordinates where the last 14 coordinates are the SD. Because the files are digitized in the same coordinate system for each face and the smoothing function performed by Pro/ENGINEER does not change orientation of the seal line file, it is possible to combine the two files. Principal components analysis is used to orient the newly combined file, which describes the entire seal area. Principal components analysis places the origin of the (x, y, z) coordinate system at the centroid of each seal area (seal line and SD) and orients the major axes along the directions of maximum variance of the seal area (30, 53). The first step is calculating the sample variance-covariance matrix $\mathbf{S}$ of the data (30):

$$\mathbf{S} = \frac{1}{n}\mathbf{A}^{\mathrm{T}}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{A} \tag{5.1}$$

where:

$\mathbf{A} = n \times 3$ matrix of (x, y, z) points of newly combined smooth file

n = number of data points

$\mathbf{I}$ = identity matrix

$\mathbf{1}$ = column vector of n ones

Next, the eigenvectors of $\mathbf{S}$ are calculated and arranged into a 3×3 matrix $\mathbf{E}$. Multiplication of $\mathbf{AE}$ transforms the matrix $\mathbf{A}$ so the axes align along the directions of maximum variance. To center the data at the centroid, the mean values of x, y, and z (of new matrix $\mathbf{AE}$) are calculated and subtracted from the coordinates of each point. The original coordinates of $\mathbf{A}$ have now been transformed so that the origin of the new coordinate system is at the center of the seal area and the major axes correspond to the direction of greatest variance.

At this point, the overall length (aligned along the x-axis) and width (aligned along the y-axis) of the seal area are determined for possible use as features for clustering purposes. The length and width are written as an (x, y, z) coordinate and appended to the end of the data file. The x-coordinate represents the length of the face, the y-coordinate the width of the face, and the z-coordinate is zero. The seal areas are now oriented to a common reference point for computational and visual convenience and are ready for the form characterization phase.

5.1.2.3 Sparsing. One of the form analysis methods considered in Section 4.7.2 requires that the data files being compared all have the same number of data points. Therefore, an additional set of data files must be created, in which each file has a fixed number of data points. Using a procedure called 'sparsing', the Pro/ENGINEER software has the ability to reduce the number of data points defining a curve to any desired number. By evaluating the seal lines of each face, it is determined that the seal line curve can be adequately represented by much fewer than the original 130 - 180 points obtained during the digitization process.

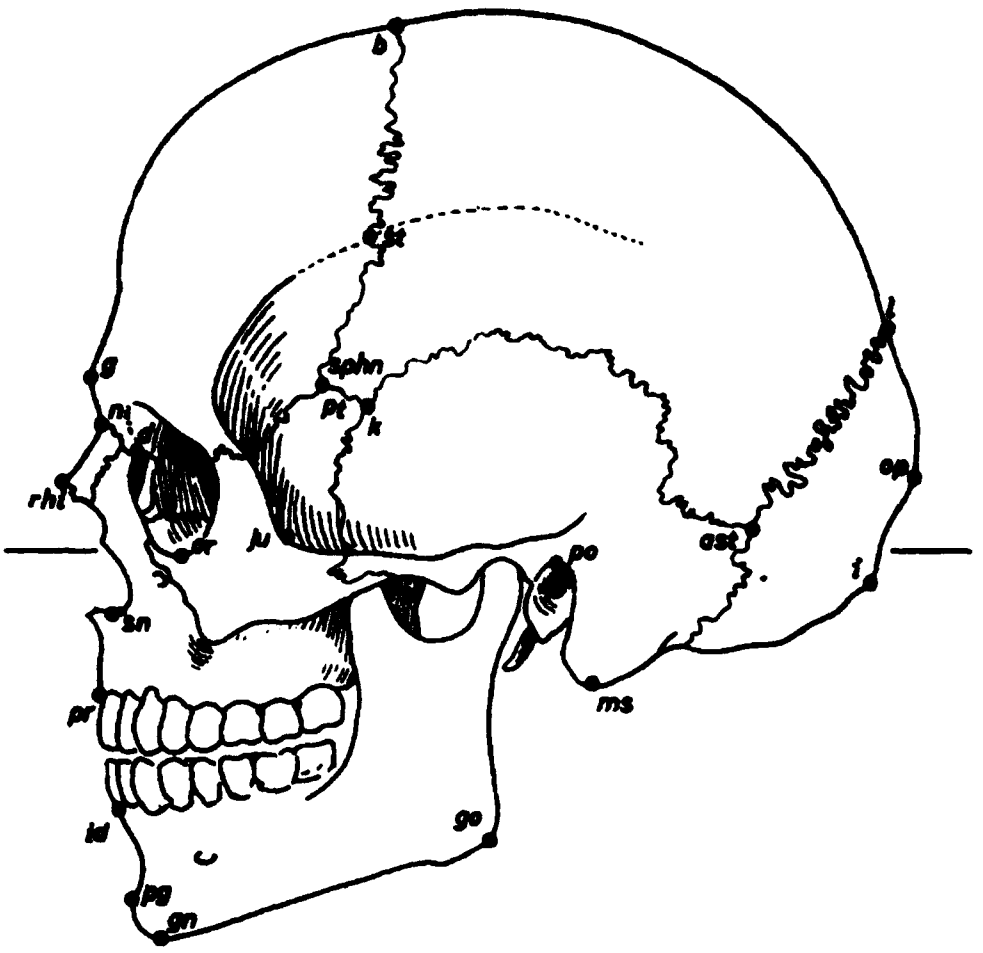


Figure 5.15 Frankfort Horizontal Skeletal Landmarks, Side View: The left *orbitale* is depicted by 'or' and the left *poria* is depicted by 'po' in the figure; these are two of the three landmarks which define the Frankfort Horizontal plane for skeletal craniometry (34).

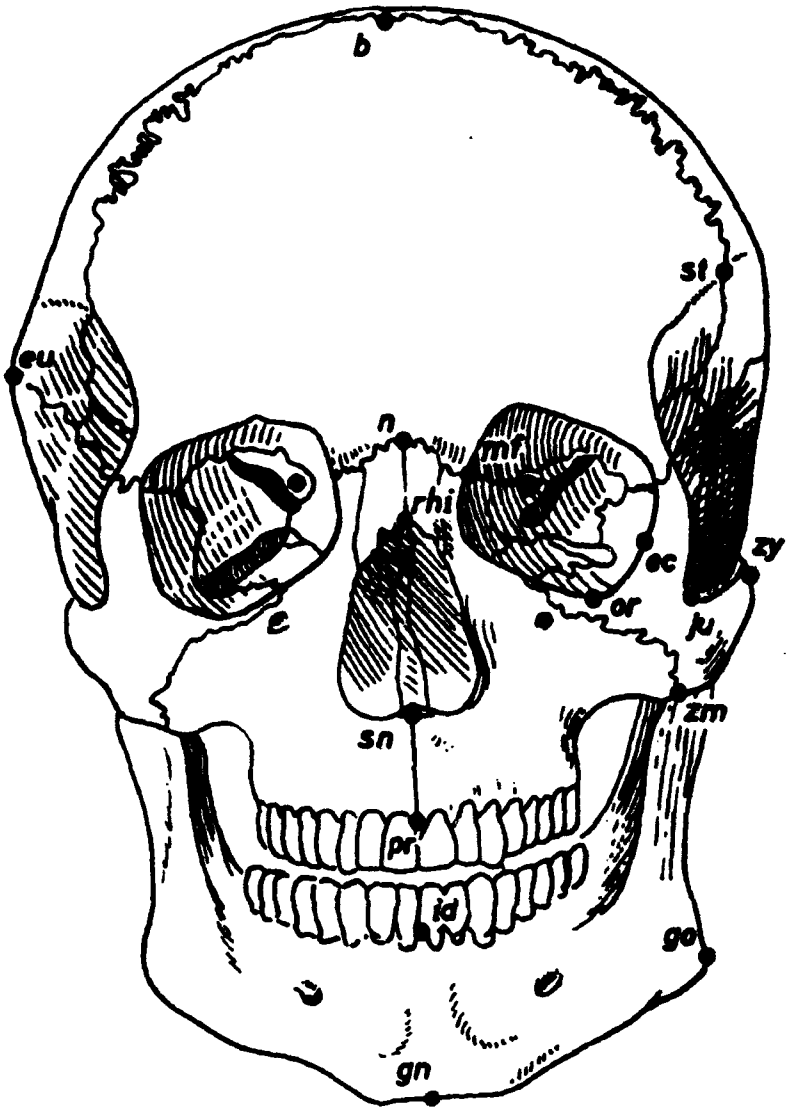


Figure 5.16 Frankfort Horizontal Skeletal *Orbitale* Landmark: The left *orbitale* is depicted by 'or' in the figure; this is only one of the three landmarks which define the Frankfort Horizontal plane for skeletal craniometry (34).

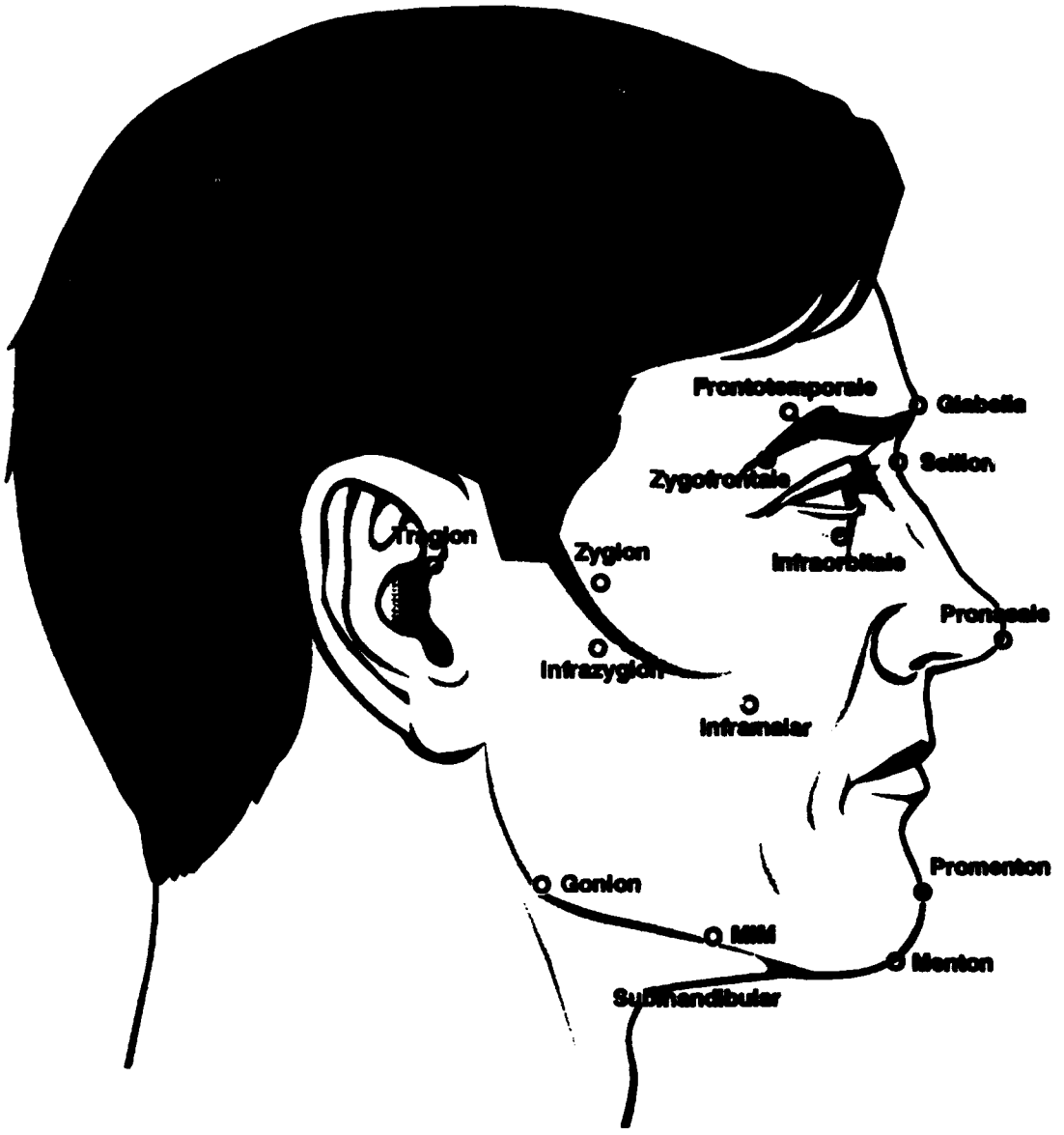


Figure 5.17 Frankfort Horizontal Live Subject Landmarks: The right *tragion* and right *infraorbitale* landmarks used to define the Frankfort Horizontal plane on live subjects.

An analysis is conducted to determine the minimum number of points needed to describe the shape of the seal line. It is not until the seal line data file is reduced to around 25 - 30 points (depending on the size of the face) that the shape of the seal line becomes significantly different. By using a smaller number of points, but still large enough to maintain the shape and size of the seal line, the data can be reduced without loss of information (see Figure 5.18). To ensure an adequate number of points are used, the seal line data file is sparsed to 39 points. To accomplish this using Pro/ENGINEER, a value is entered that sets the maximum allowable deviation from the original curve. The sparsing program deletes points on the curve and calculates a new curve passing through the remaining points. As many points as possible are deleted, without allowing the new curve to differ from the original curve by more than the deviation value. It is found that by selecting a deviation in the range of 0.3mm to 0.4mm, the seal line files can be reduced to 39 data points. This reduction in the number of points results in an insignificant change in the shape of the seal line curves calculated by Pro/ENGINEER, while reducing the size of each seal line data file to a consistent number across the population.

5.1.3 Data Collection and Preparation Summary. Plaster face casts are obtained from the CMS with seal lines drawn by lab technicians. The seal lines and 14 spacial delineators (SD) are digitized using a three-dimensional digitizer. The seal line and the 14 SD are saved as two separate files, but are digitized in the same coordinate system. Combined, these two files represent the 'seal area' for each face. During the data preparation phase, two new files per face are created for use in form analysis. The first file contains the smoothed seal line (obtained through Pro/ENGINEER), the 14 SD, and the length and width of the seal area. These files contain anywhere from 130 to 180 data points. The second file contains only 54 points, consisting of the 39 sparsed seal line curvature points, 14 SD, and the (x, y, z) coordinate describing the length and width of the face. The next step is characterizing the form of the seal areas.

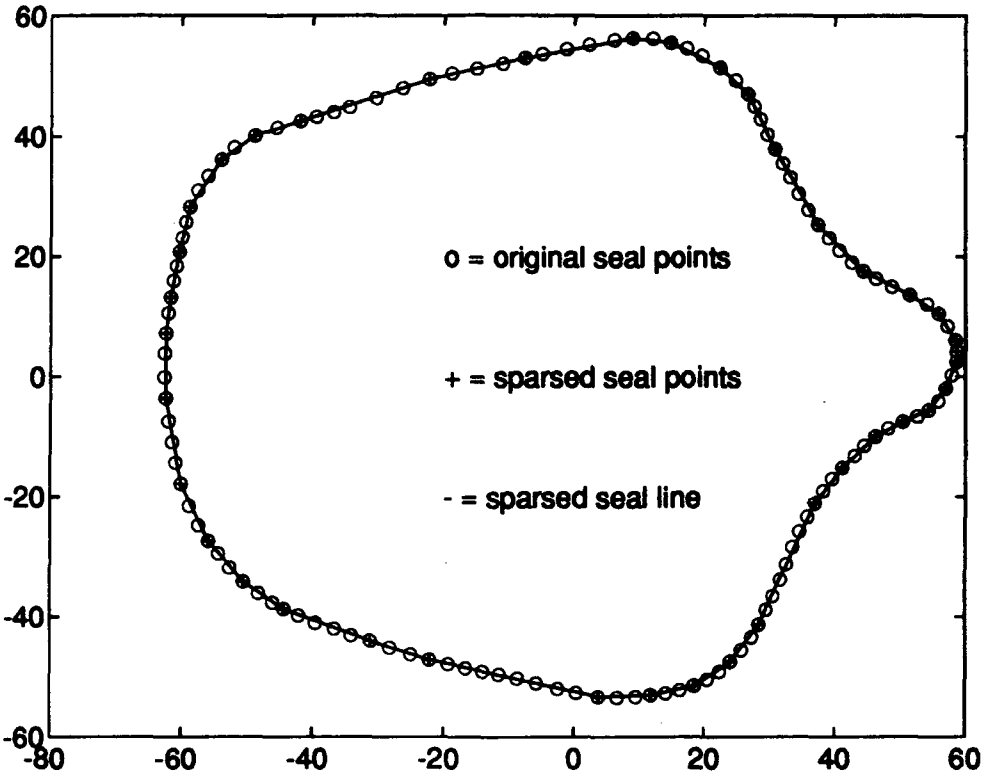


Figure 5.18 Sparsed Curve: The line is a curve calculated using ONLY the 39 points representing the sparsed data set. Notice that by using only the sparsed data points, the curve does not deviate from the original points.

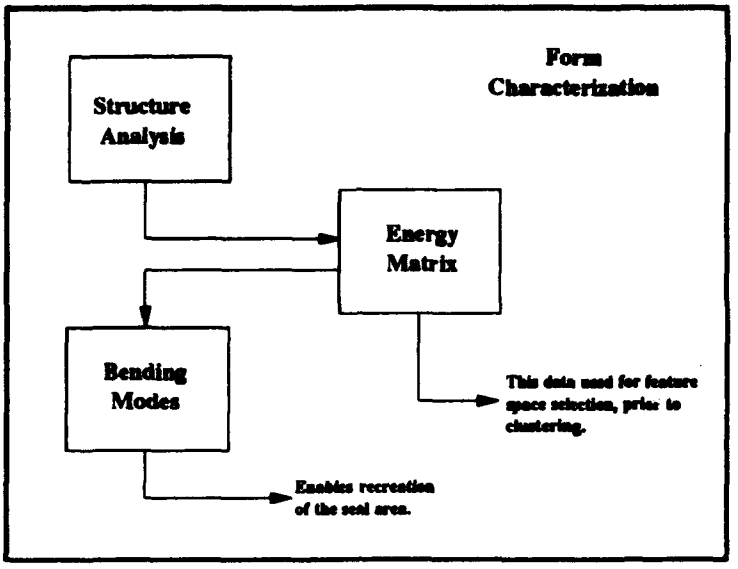


Figure 5.19 Form Characterization: Structure analysis development of energy matrix.

5.2 Form Characterization

This section derives the spacial correlation for the seal areas. The structure analysis on each face results in selection of the spherical variogram model (discussed in Section 5.2.1) to represent the spacial correlation for the human face. The spherical variogram model provides the information to build the energy matrix described in Section 5.2.2.

5.2.1 Structure Analysis. The process of studying spacial correlation and developing a variogram function is referred to as structure analysis. Spacial correlation is the correlation of data points separated by a given distance and is inversely related to the variogram function. Points close together are highly correlated (small variogram) since knowledge at one point is a good indication of the value of nearby points. Conversely, points far apart have low correlation (high variogram) since knowledge of one point gives very little information about a point a large distance away. At sufficiently large distances the variogram is equal to the overall variance of the data. The distance at which the variance of the data is achieved is defined as

the range of influence of the data. There are several steps involved in the structure analysis of the data.

The first step in structure analysis is to calculate the correlation of the data as a function of the separation between points. The correlation is used to calculate an experimental variogram. The experimental variogram for the seal areas show that an overall trend or drift exists which is removed using residual analysis. The experimental variogram on the residual exhibits weak stationarity (Section 3.2.1).

The next step is to fit a functional variogram to the experimental variogram. The functional variogram is a continuous function, while the experimental variogram is defined only at discrete points (i.e., the separation distance of the data). Of the several existing functional variogram models described in Section 3.2.1, the spherical model is chosen as the best fit to the facial data. Least squares is used to estimate parameters of the functional model. The functional model also incorporates the presence of geometric anisotropy in the data.

The result is a trend free, isotropic, functional variogram which depends only on the variance in height, z, of the face and the separation distance, h, of the points being compared.

5.2.1.1 Experimental Variogram. The first step in structure analysis is to calculate the correlation structure and experimental variogram of the data. The experimental variogram is estimated using the following equation (Section 3.2.1.2):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{n} [z(\mathbf{x}_i + \mathbf{h}) - z(\mathbf{x}_i)]^2 \qquad \text{for } 0 \leq h \leq m \tag{5.2}$$

Strictly speaking, this variogram is calculated by finding the variance of all points exactly 1, 2, ..., or m units apart in all directions, where m is some distance greater than or equal to the the major axis or largest distance. The distance h is calculated as a function of (x, y), and the variance is based on the height, $z(x, y)$, at each

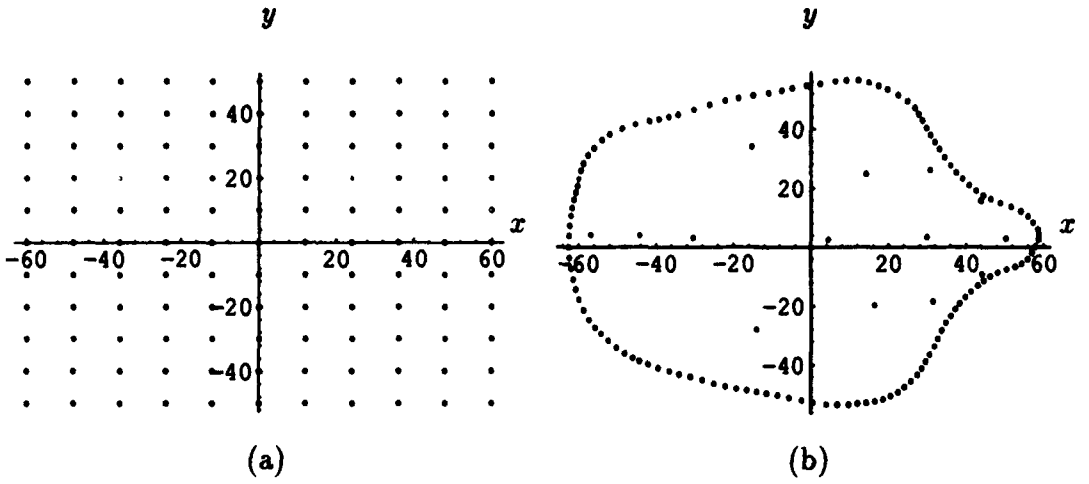


Figure 5.20 Possible Kriging Grid Shapes: (a) Perfect grid with points separated exactly 12 units in x direction. (b) Imperfect grid of facial points separated by irregular distances.

location. This can be done only if the data is distributed in a perfect grid as shown in Figure 5.20a. There are cases when the data can be taken in a perfect grid. For example, a mining engineer often has the control to take ore samples in any interval or grid desired. However, the seal area data of an oxygen mask is not oriented in a perfect grid as shown in Figure 5.20b.

Because of the irregularity of facial data, the correlation can not be calculated for *exact* separation distances. Instead, the correlation is analyzed at separation *ranges* of 0-1 mm, 1-2 mm, etc. By incrementing h in steps of one millimeter, the experimental variogram is calculated for all separation distances and plotted in Figure 5.21a. This figure is the experimental variogram of the raw data. There are two irregularities in the experimental variogram.

The first problem with the experimental variogram occurs at distances of about 60 to 100 mm. At these distances the variogram is greater than the variance of the data. Physically, this means that estimation of heights at unknown points located 60 to 100 mm from a known point is 'less' accurate, or has less correlation, than

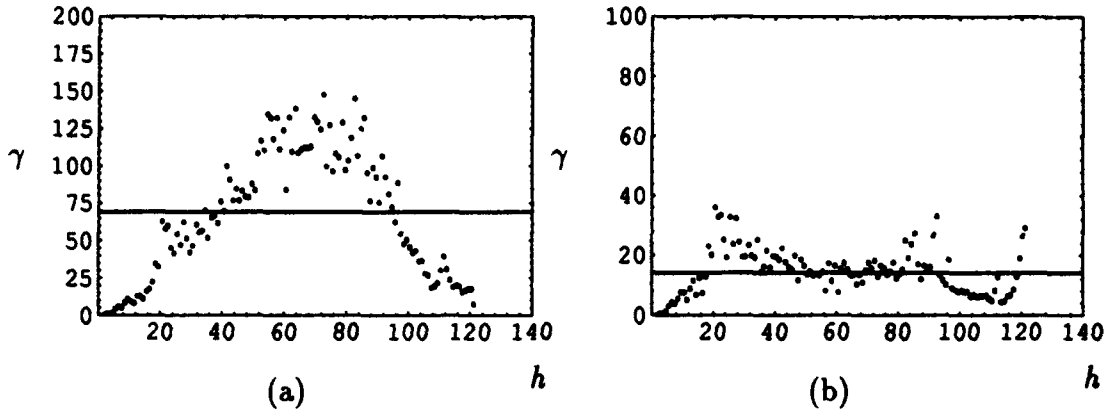


Figure 5.21 Experimental Variogram of Seal Area Data: Experimental variogram vs separation distance for (a) original data with variance of 69.0 and (b) residual data with variance of 14.1.

a random guess based on the variance of the data. It is undesirable to have the variogram exceed the variance of the data.

The second problem with the experimental variogram occurs at distances of about 100 mm and greater. At these distances the variogram drops 'below' the variance of the data. Points separated by relatively large distances should have no correlation and the variogram should be equal to the variance of the data. However, in this case it appears that at large distances the correlation increases. These problems signify the presence of a trend in the data due to the symmetry of the human face, realizing that at large distances (across the breadth of the face) the height of the face is very similar.

5.2.1.2 Removing Trend in Data. There are two problems with the experimental variogram of the seal area data. As can be seen in Figure 5.21, it rises above the sample variance of the data over the range of 60 to 100 mm, and is below the sample variance for distances greater than 100 mm. These problems are due to an overall trend in the data. The trend is removed by calculating the residuals.

The residuals exhibit weak stationarity, and are used to calculate the experimental variogram (19).

The first step in the residual analysis is to fit a polynomial through the oriented facial data (see Section 3.2.3). Choosing a second order polynomial fit of the data in order to approximate the trend, $z'(x, y)$, gives:

$$z'(x, y) = b_1 + b_2 x + b_3 y + b_4 x^2 + b_5 y^2 + b_6 xy \tag{5.3}$$

where the coefficients vector **b** is different for each face. The polynomial surface is a least squares regression of the data and does not pass exactly through all the data points, $z(x, y)$. The residual, $e(x, y)$, is the difference between the actual measured data point and the trend at each location:

$$e(x, y) = z(x, y) - z'(x, y) \tag{5.4}$$

5.2.1.3 Experimental Variogram of Residual Data. The third step in the structure analysis is to calculate the experimental variogram of the residual data. If the experimental variogram is trend free, then the second order polynomial represents a good fit of the data. If the variogram still has a trend, a higher order polynomial fit of the data is required.

The result of calculating the experimental variogram (see Section 5.2.1.1) of the residual data is shown in Figure 5.21b. The experimental variograms of the residuals of the seal areas rise and stay at the variance of the data, showing that the trend has been removed.

5.2.1.4 Functional Variogram Model. The final step in the structure analysis is to fit a continuous or functional variogram model to the experimental variogram plots of the residuals. The functional variogram model is defined at all

separation distances, not just at discrete points. Also, this continuous variogram model must satisfy the three assumptions for a variogram function (Section 3.2.1.1).

The first assumption of a variogram function is that the variable being modeled is regionalized. A regionalized variable has continuity from point to point, but changes in ways that are too complex to be described by a deterministic function. The experimental variogram is an increasing function at small separation distances, which indicates continuity from point to point. However, at larger separation distances, the variogram levels out, and estimating the regionalized variable is more difficult. Therefore, the experimental variogram satisfies the first assumption of a variogram function.

The second assumption required is that the data being modeled exhibit at least weak stationarity (Section 3.2.1.1). Data exhibit weak stationarity if the first and second moments (mean and variance) are not functions of location on the face. The raw data do not satisfy this requirement. The residuals, however, do exhibit weak stationarity, and so are used in all follow-on analysis.

The final assumption is that the data being modeled are spacially isotropic. The data are assumed isotropic if the sample variance and range of influence do not change with the orientation of **h**, the vector between two points. If geometric anisotropy is detected, then a correction factor can be found which restores isotropy to the variogram. The seal areas do exhibit anisotropy, which is corrected for in order to satisfy this assumption.

5.2.1.5 Fitting Functional Model. A functional variogram is determined by selecting a simple model which best fits the experimental variogram. Several models are tried, and the best fit to the experimental variogram of the residuals

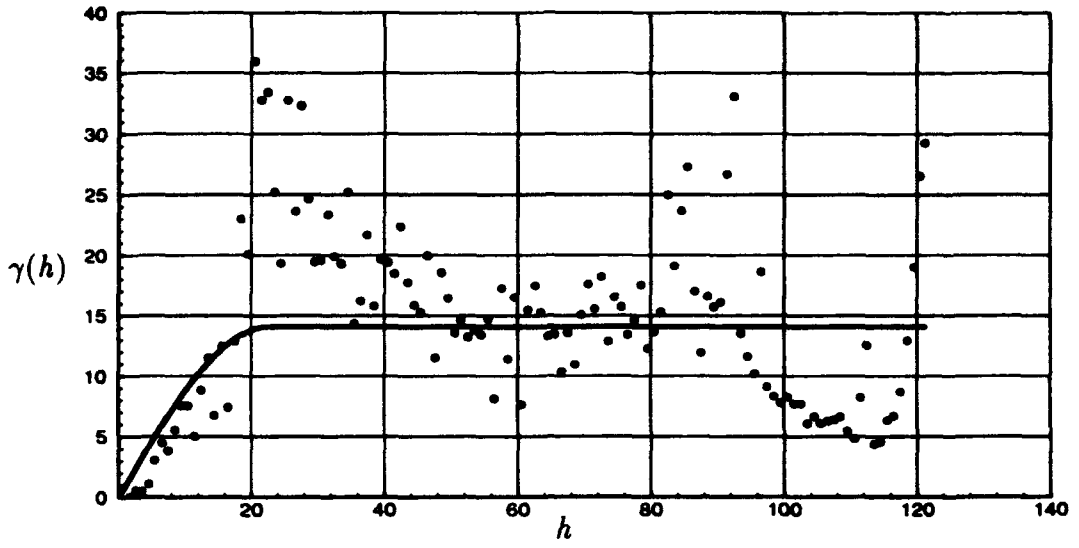


Figure 5.22 Spherical Variogram of Residual Data: Functional variogram vs separation distance for residual data fitted with spherical model.

is the spherical model (Figure 5.22):

$$\gamma(h) = \begin{cases} 0 & \text{when } h = 0 \\ Co + C\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) & \text{when } 0 < h < a \\ Co + C & \text{when } h \geq a \end{cases} \tag{5.5}$$

where h is the separation distance $\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. The unknown parameters in these models are the nugget effect Co, the range of influence a, and the sill $Co + C$, which must be calculated for each face.

The nugget effect is the variance of data separated by very small distances. For continuous data the nugget effect is zero since points infinitesimally close together are perfectly correlated to each other. Conversely, for white Gaussian noise the nugget effect is equal to the variance of the data, since points infinitesimally close together have no correlation to each other. When estimating ore content, the variance of nearby samples may not be zero due to nuggets of ore creating discontinuities in the ore levels, hence the term 'nugget effect'. When estimating the shape of the face,

the nugget effect may be nonzero due to measurement errors. However, errors in the digitization of the seal areas are corrected during data preparation so that $Co \equiv 0$.

The sill is the overall sample variance for a seal area. This is estimated using the calculations for the experimental variogram of the residuals e at large separation distances, or:

$$\sigma^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (e - \bar{e})^2 \tag{5.6}$$

Since the nugget effect is zero, the sill $(Co + C)$ reduces to C, and is the variance in the residuals, e, of the face.

The range of influence is the separation distance at which the variogram attains the sill of the data. The range of influence is calculated by minimizing least squares error in the fit of the functional variogram model to the experimental variogram. Many computer codes have been written to automatically calculate the variogram parameters; see, for example (27).

The spherical model provides the best fit to the data at small separation distances, and gives the most consistent results. Therefore, for the purpose of this research, the spherical model is chosen to represent the variogram of the residuals on the seal areas.

5.2.1.6 Anisotropy. One assumption required in the use of experimental variograms is that the data are isotropic. This is checked by calculating correlations in different directions. The previous experimental and functional variograms are calculated regardless of direction. Recall that since $\mathbf{h}$ is a vector, it has a specific orientation, $\mathbf{h} = (x_i - x_j, y_i - y_j)$. This orientation becomes important when dealing with anisotropic behavior. If the experimental variogram is dependent on direction, the data are anisotropic. There are two types of anisotropy: geometric and zonal (see Section 3.2.2). The next step is to calculate the experimental variogram in several directions and fit the spherical model to each experimental variogram. The type is determined by finding the parameters, a and $Co + C$, of the functional

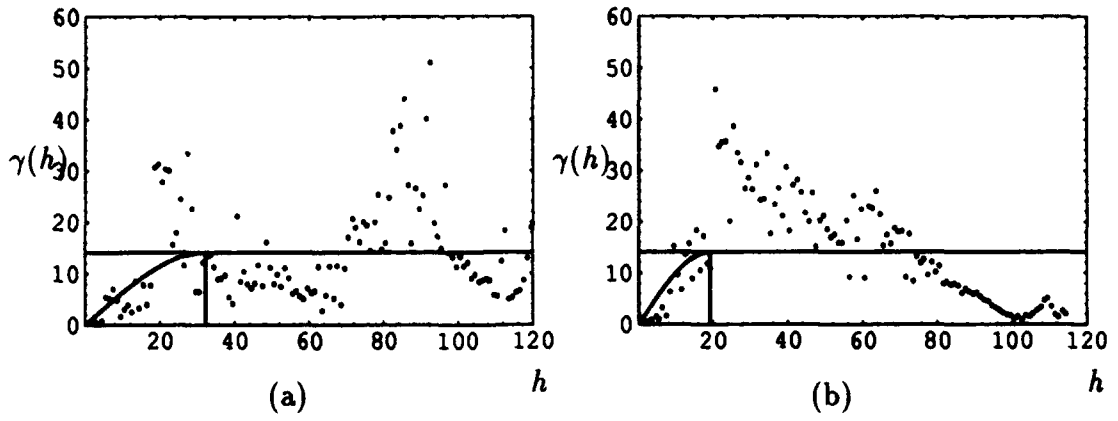


Figure 5.23 Variograms of Anisotropic Residual Data: Functional variogram vs separation distance for residual data in two different directions.

variogram in several directions. The data are isotropic if a and $Co+C$ are the same in all directions. Figure 5.23 shows the spherical variogram in different directions for a single face. This variogram plot shows that while the sill remains the same, the range of influence of the data changes with direction. This indicates the seal area data are geometrically anisotropic.

The simplest method to correct for geometric anisotropy is to scale the principal axes of the data. The anisotropic correction factor is found by calculating the a's for various directions. Figure 5.24 is obtained by calculating the a for directions starting at 10°, and incrementing by 20°, up to 180°, with a narrow 'window' of 20°. The window sets limits on the orientation of the distances to be evaluated. In this case, at 10°, the window spans from 0° to 20°, and only evaluates orientations of h that fall in this range, for a given length. The directions only need to cover 180°, since $\gamma(h) = \gamma(-h)$. Figure 5.24 enables the identification of the major and minor axes of the range of influence (29). The ranges of influence (or rose diagram) typically represents an ellipsoid, where the major and minor axes are also known as the maximum and minimum continuities. The rose is not a perfect ellipse due to

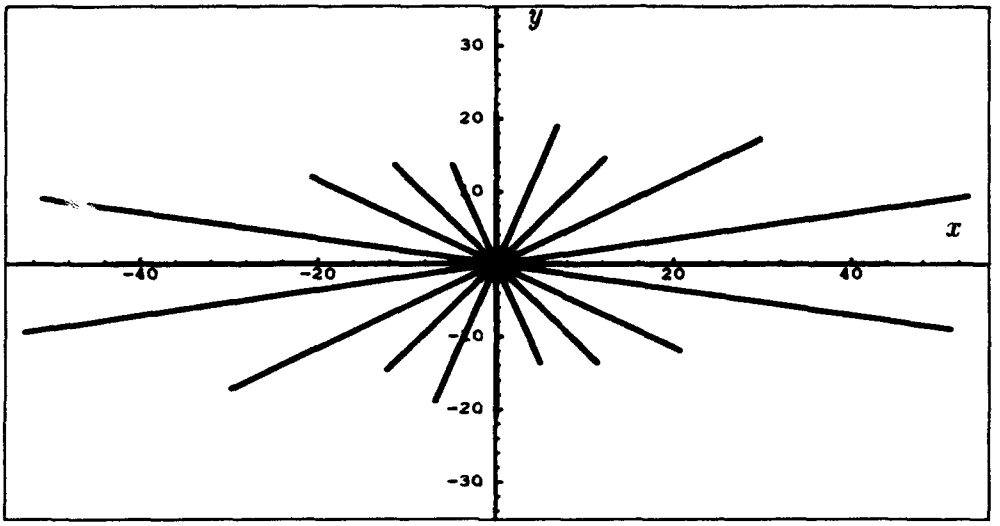


Figure 5.24 Anisotropy Rose: The range of influence, a as a function of direction.

some asymmetry in the seal area, especially along the x-axis. By fitting an ellipse to the figure, the maximum and minimum ranges of influence are aligned in the x and y directions. This is because the data are already oriented in principal components. The correction factor is defined by:

$$k = \frac{a_{\text{maj}}}{a_{\text{min}}} \tag{5.7}$$

which is incorporated into the (x, y) Euclidean distance calculation. The range of influence in the x direction is $a_x = a_{\text{maj}}$ and the range of influence in the y direction is $a_y = a_{\text{min}}$. Applying this correction, the magnitude of $\mathbf{h}$, or the separation distance h is:

$$h_{ij} = \sqrt{(x_i - x_j)^2 + [k(y_i - y_j)]^2} \tag{5.8}$$

Therefore, since the faces are already oriented in this research, the major and minor a's are calculated using directions along the x and y axes or 0° and 90°, with a window of 90°. This range is shown in the double arc in Figure 5.25. The figure shows that points 14, 15, 16, 17, 18, and 19 are separated from point 138 by 70 to 71 mm in the x direction. By moving the arc to each data point there are a total

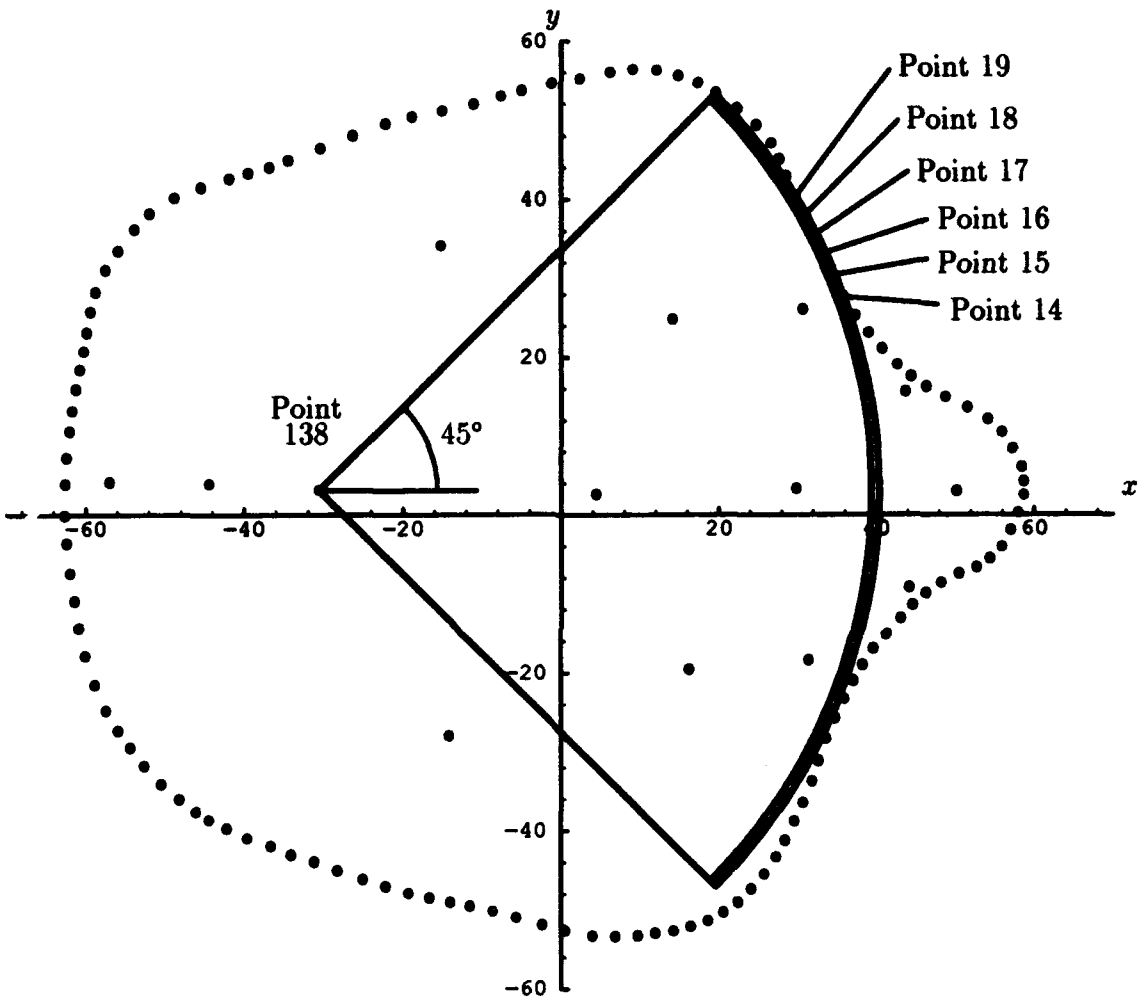


Figure 5.25 Variogram Estimation Arc: Arc showing points 14, 15, 16, 17, 18, and 19 are separated from point 138 by a distance of 70 to 71 nm in the x direction. The arc is a 90° window pointing in the positive x direction, or 0°.

of 36 points separated by 70 to 71 mm in the x direction. The arc looks only in the positive x direction to avoid counting points twice.

5.2.1.7 Constant Range of Influence. The final step in calculating the functional variogram model is to find the average range of influence of the human face. Based on the best fit of the experimental variogram function, the range of influence for each face varies slightly. However, all human faces are similar to each other, so the range of influence of any two faces should be approximately the same. To correct for this, a_x and a_y are averaged for the 99 faces in the data set. The mean values for the ranges of influence are $a_x = 33.621$ and $a_y = 17.655$, resulting in a correction factor of $k = 1.90$.

5.2.1.8 Summary of Structure Analysis. This section has explained how the structure analysis of the data is used to calculate the functional variogram model. The residuals of the data are calculated to remove the trend in the data. Several different models are fit to the experimental variogram, and it is found that the spherical model produces the best match to the seal area data. Anisotropy is identified, and a correction factor, k, is derived and incorporated into the distance calculation. The spherical variogram includes the range of influence for the 'average' human face and the variance of each individual face. Substituting the range of influence for the major axis, the modified separation distance, and zero nugget effect into the spherical variogram model gives:

$$\gamma(h) = \begin{cases} 0 & \text{when } h = 0 \\ C\left[\frac{3h}{22.414} - \frac{h^3}{76{,}008}\right] & \text{when } 0 < h < 33.621 \\ C & \text{when } h \geq 33.621 \end{cases} \tag{5.9}$$

where C is the variance of each individual face. The resulting functional variogram model is a function of the overall variance of a face and the separation distance of the points being compared.

5.2.2 Energy Matrix. The energy matrix provides a unique form characterization for a face. This section describes how the energy matrix is constructed using the variogram calculated in the previous section, and discusses how the eigensystem decomposition of the matrix can be viewed as bending modes of the face.

5.2.2.1 Steps to Create $\mathbf{L}_n^{-1}$. The structure analysis done in the previous section was based on the unsparsed data. This resulted in removal of the trend, selection of the spherical variogram model, and determination of the average ranges of influence over all faces. However, the most consistent clustering results came when using the sparsed data. For this reason, the energy matrix is created using the residuals of the sparsed data, while still using the variogram parameters (C , $a_{\mathbf{maj}}$, and k) from the unsparsed data set.

Step 1: Calculate the distance between each pair of data points,

$$h_{ij} = \sqrt{(x_i - x_j)^2 + [1.90(y_i - y_j)]^2} \tag{5.10}$$

where $i = 1, ..., n$ and $j = 1, .., n$, with n being the total number of data points. Distance h_{ij} represents the Euclidean distance in the xy plane only, and includes the anisotropic correction factor, $k = \frac{a_{\mathbf{maj}}}{a_{\mathbf{min}}} = 1.90$ ($a_{\mathbf{maj}}$ occurs in the x-direction in all of the seal areas). The distances form the $\mathbf{H}$ matrix:

$$\mathbf{H} = \begin{bmatrix} 0 & h_{12} & h_{13} & \dots & h_{1j} \\ h_{21} & 0 & h_{23} & \dots & h_{2j} \\ h_{31} & h_{32} & 0 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & h_{n-1,n} \\ h_{i1} & h_{i2} & \dots & h_{n,n-1} & 0 \end{bmatrix} \tag{5.11}$$

Note that the matrix $\mathbf{H}$ is identical to the two-dimensional EDMA matrix, $\mathbf{F}(\mathbf{A})$ in Equation 2.3.

Step 2: Calculate the variogram $\gamma(h_{ij})$ using the spherical model, where the parameters for each face, $Co = 0$, $a_{\text{maj}} = 33.621$, and C are derived from structure analysis:

$$\gamma(h) = \begin{cases} 0 & \text{when } h = 0 \\ C\left[\frac{3h}{22.414} - \frac{h^3}{76{,}008}\right] & \text{when } 0 < h < 33.621 \\ C & \text{when } h \geq 33.621 \end{cases} \tag{5.12}$$

The function $\gamma(h)$ is applied to each element of **H** resulting in a new matrix **K** (as in Equation 3.73):

$$\mathbf{K} = \begin{bmatrix} 0 & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) \\ \gamma(h_{21}) & 0 & \cdots & \gamma(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & 0 \end{bmatrix} \tag{5.13}$$

The variogram, $\gamma(h_{ij})$, is now a function only of the scalar distance h_{ij} from (x_i, y_i) to (x_j, y_j). Two seal areas with similar variances have different energy matrices since the separation distance h_{ij} varies from face to face, depending on size and shape.

Step3: Create the **L** matrix. In both universal kriging and TPS, the **L** matrix includes the localized trend analysis, Equation 3.72, where (x_i, y_i) are the coordinates of the known data points, and n is the total number of data points (Section 3.2.4.2): By inserting the spherical variogram model into the universal kriging matrix, the following **L** matrix is developed:

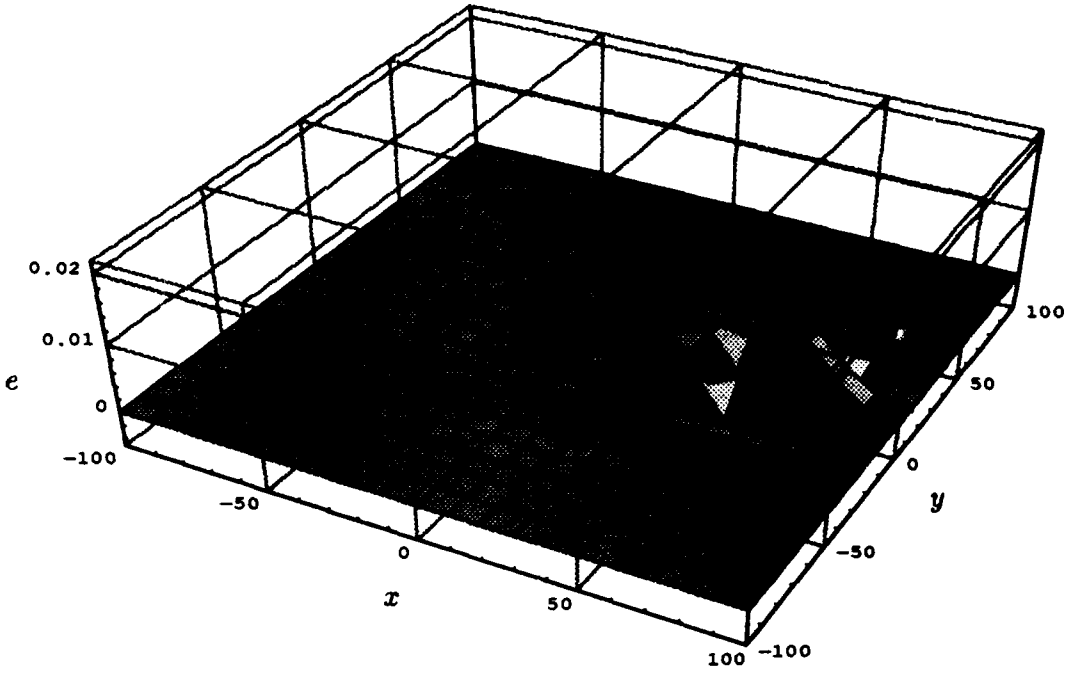
$$\mathbf{L} = \begin{bmatrix} \nwarrow & \uparrow & \nearrow & 1 & x_1 & y_1 \\ \leftarrow & \mathbf{K} & \rightarrow & \vdots & \vdots & \vdots \\ \swarrow & \downarrow & \searrow & 1 & x_n & y_n \\ 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & \cdots & x_n & 0 & 0 & 0 \\ y_1 & \cdots & y_n & 0 & 0 & 0 \end{bmatrix} \tag{5.14}$$

Step 4: Calculate $\mathbf{L}^{-1}$, and create $\mathbf{L}_n^{-1}$. Since the $\mathbf{L}$ matrix is well-conditioned (real and symmetric), the inverse exists and is unique to each individual face (29). The entire $\mathbf{L}$ matrix is inverted, and the upper left $n \times n$ matrix, $\mathbf{L}_n^{-1}$, is formed by merely dropping the last three rows and columns. The energy matrix is now analyzed to determine the form of the face it represents.

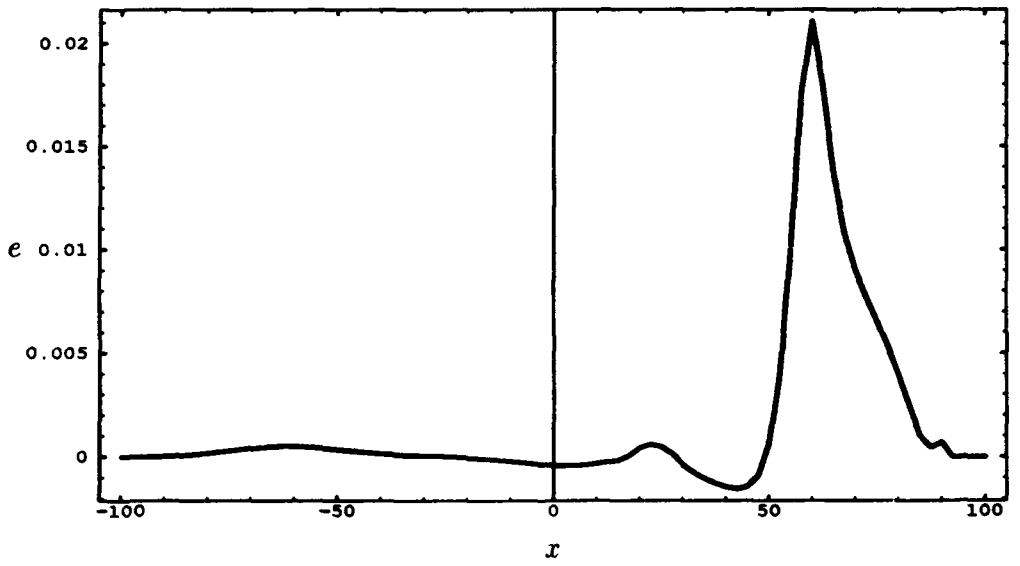
5.2.2.2 Bending Modes of Facial Data. As discussed in Section 3.4.2, one possible method of analyzing the energy matrix is via the eigensystem decomposition. The eigenvalues and eigenvectors can be viewed as individual bending modes of a face. Figures 5.26a and 5.27a show the bending modes for the largest and smallest eigenvalue of $\mathbf{L}_n^{-1}$ for mask identification number 0002. Figures 5.26b and 5.27b are cross sections of the respective bending modes along the plane $y = 0$ (i.e., the vertical axis of the face). The vertical axis is the residual, e, which is being predicted. These show (as in the bending mode example section 3.4.2.2) that large eigenvalues of $\mathbf{L}_n^{-1}$ relate to localized bending and small eigenvalues relate to global bending. The eigenvectors show the location on the face where the bending is located. The overall magnitude of the bending, as shown by the different scales, is determined largely from the mode constant $\psi = \mathbf{e}^{\mathrm{T}}\mathbf{z}$.

The sum of the fifty bending modes and the regression plane is the residuals of the faces. Finally, the residuals plus the trend surface results in the predicted physical height of the face (Figure 5.28). Appendix A contains the complete set of fifty bending modes and shows the recombination into physical facial dimensions.

The energy matrix contains information that uniquely characterizes the physical form of a face. In order to group faces into similar clusters, the information contained in the energy matrix must be put into a form compatible with clustering. Since no existing clustering routine can operate on matrices, the energy matrices must be reduced in some way. During this reduction, the vital form characteristics

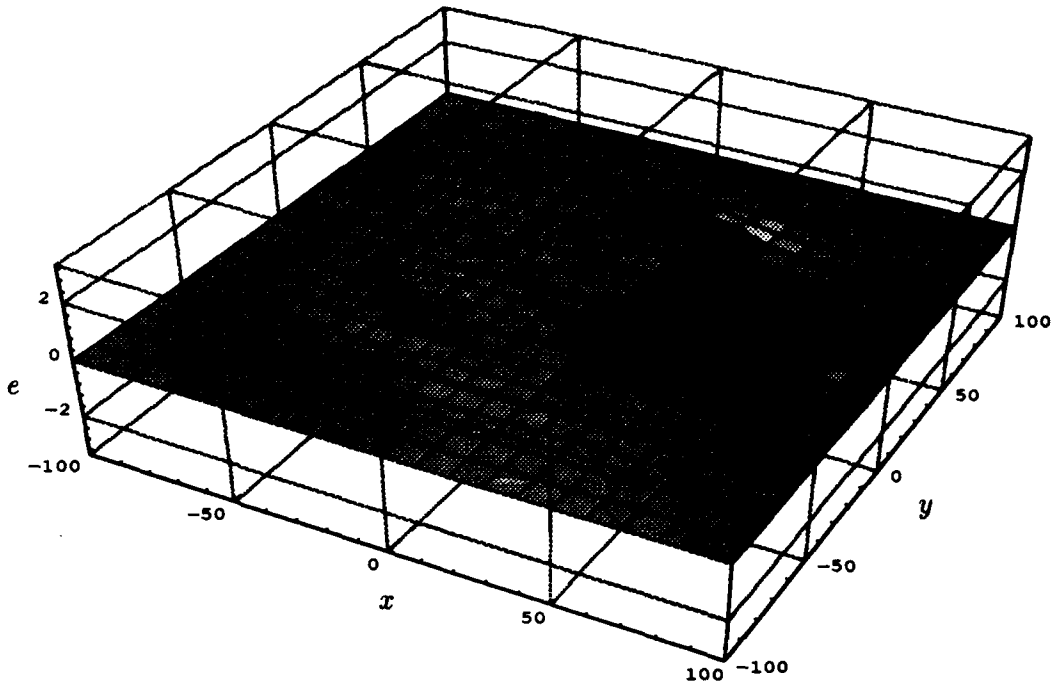


(a) Full Bending Mode

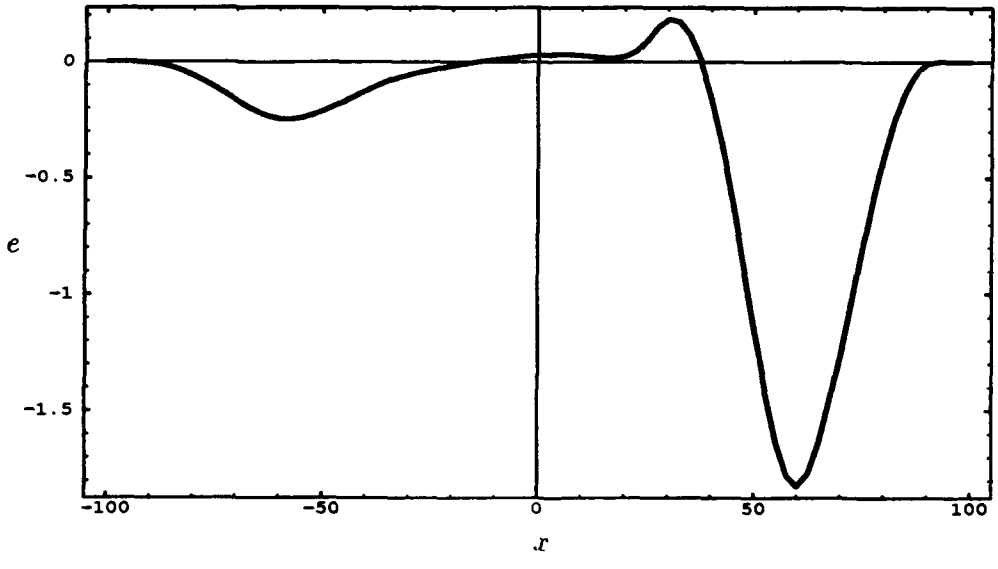


(b) Cross-section at $y = 0$

Figure 5.26 Facial Bending Mode 1: $\lambda_1 = -0.73799$.

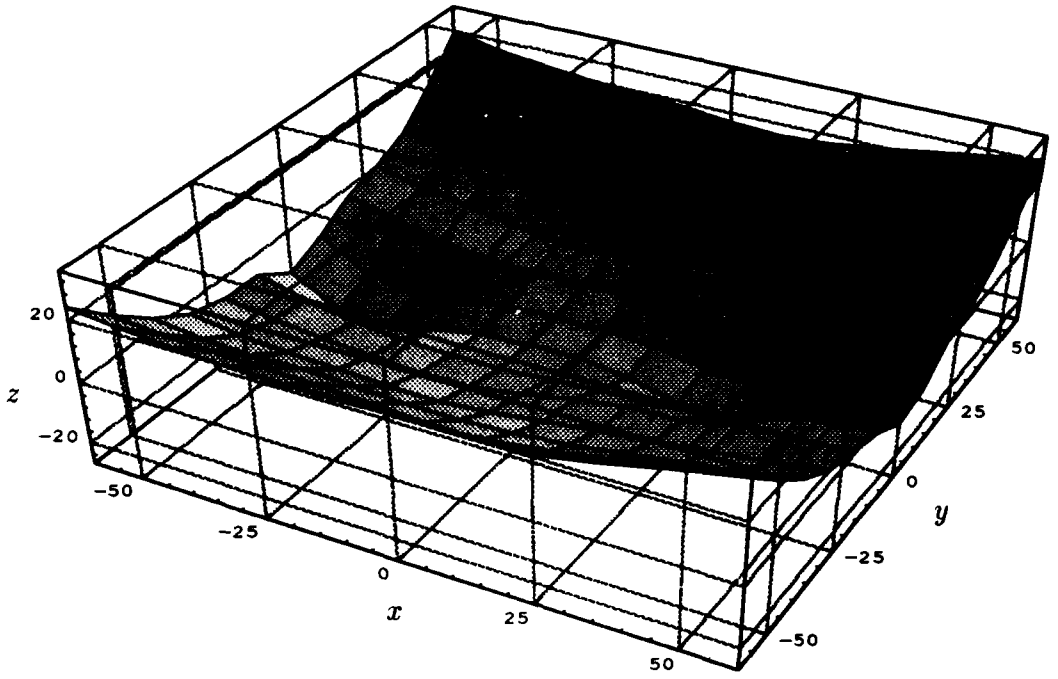


(a) Full Bending Mode

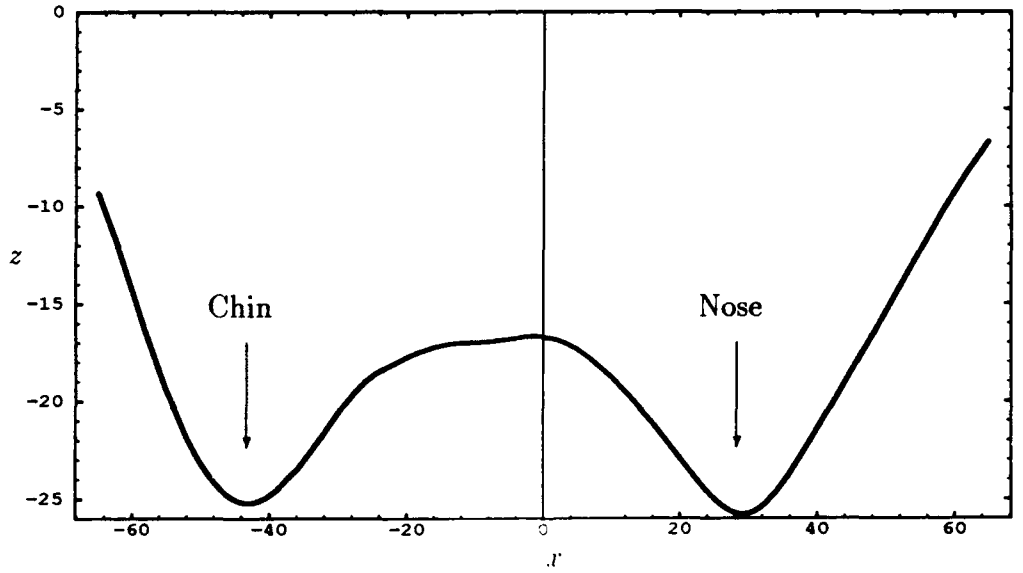


(b) Cross-section at $y = 0$

Figure 5.27 Facial Bending Mode 50: $\lambda_{50} = -0.021602$.



(a) Full Bending Mode



(b) Cross-section at $y = 0$

Figure 5.28 Face Reconstruction: Estimated hight of face in physical dimensions.

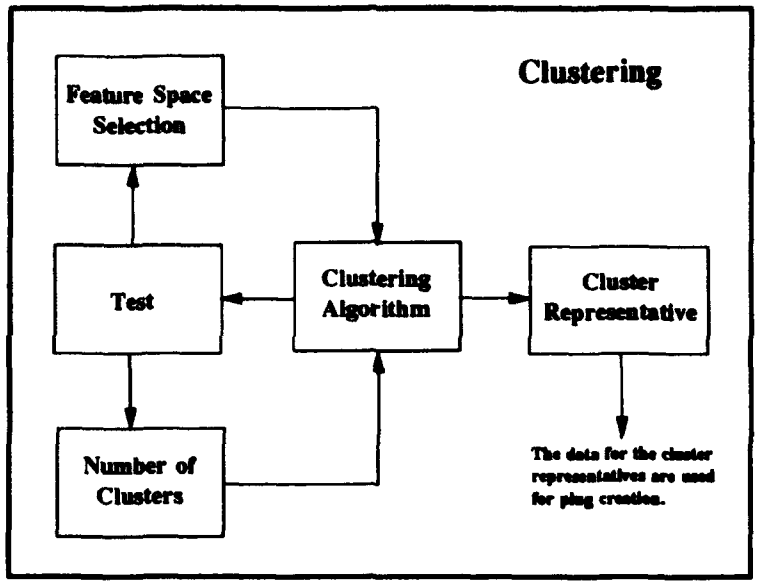


Figure 5.29 Clustering.

of the seal area must be retained, while unimportant information is eliminated. This stage is called feature space determination.

5.3 Clustering

The six different feature sets discussed in Chapter IV are used in the fuzzy c-means algorithm. Based upon the unique descriptive nature of the $\mathbf{L}_n^{-1}$ matrix, these features appear to hold the most promise for clustering the seemingly unclusterable faces. To review, the feature sets compared are:

- every 20^{th} percentile eigenvalue of the unsparsed $\mathbf{L}_n^{-1}$ matrix
- the top 5 eigenvalues of the unsparsed $\mathbf{L}_n^{-1}$ matrix
- the top 5, inverted, nonzero eigenvalues of the unsparsed $\mathbf{L}_n^{-1}$ matrix
- every 20^{th} percentile inverted, nonzero eigenvalue of the sparsed $\mathbf{L}_n^{-1}$ matrix
- the KLT coefficients of the eigenvalues of the sparsed $\mathbf{L}_n^{-1}$ matrix
- the KLT coefficients of the sparsed $\mathbf{L}_n^{-1}$ matrix

Also recall that the clustering results are based upon the following:

1. The number of clusters is less than or equal to six. This value is determined using engineering judgement outlined in Section 4.6.
2. The maximum number of features is restricted to five. The decision to use five features is an engineering decision based upon literature pertaining to dimensionality constraints (15, 21, 32, 44, 54).
3. The clustering is accomplished using the fuzzy c-means algorithm outlined in Section 4.5.1.

A sample is considered to belong to a cluster if it achieves a membership of at least 0.80 in that cluster. Any sample not achieving at least 0.80 membership in any cluster is considered indeterminant. This threshold membership value is set from practical experience in the use of the fuzzy c-means algorithm.

5.3.1 Clustering Results. From these six clusterings, a decision is made on which set of features best represents the actual distribution of the faces. This is done by examining the 99 plaster face casts, grouping faces according to the different clusterings, and visually estimating if the cluster memberships actually indicate similarity between faces. The results of this empirical study show the fixed form KLT on the entire energy matrix performs the best. In this feature space, faces with memberships above 0.90 have many similarities in size and shape which are not seen in the other clusterings. The KLT on the eigenvalues of the energy matrix also performs well, but has some inconsistencies which do not appear in the KLT of the entire matrix. It is hypothesized that more information is lost in taking the KLT on the eigenvalues of the energy matrix, than in performing the KLT directly on the fixed form energy matrix.

A second validation of the feature set selection also supports the selection of the KLT on the sparsed energy matrix. Face 1963 has two separate digitizations (1963 and 1963-2). Each digitization is slightly different in the location of the spacial delin-

eators and the number of points recorded around the seal curve. The purpose behind the existence of two matrices for the same face is twofold: to test the robustness of the energy methods to variations in data collection and to test the validity of the feature space selection. If the chosen feature set truly represents facial forms, 1963 and 1963-2 should end up in the same cluster with similar memberships. Using the KLT on the $\mathbf{L}_n^{-1}$ energy matrix, 1963-2 has a membership of 0.90 and 1963 has a membership of 0.86, for a difference of 0.04. This difference is by far the smallest of all the feature sets attempted in this research.

The results from clustering on this feature space are presented in Figure 5.30. The figure shows the hierarchical grouping of the faces as the number of clusters increases from two to six. The numbers in the arrow head blocks on the left represent the number of clusters input into the clustering routine. Each row shows the results for that number of clusters, using five features. Therefore, the top row has two blocks and the last or bottom row has six blocks. Each block contains the identification numbers corresponding to the plaster face casts. As the number of clusters increases, some faces consistently maintain high memberships (above 0.80) in the clusters, which allow for the hierarchical organization. On the fringe, other faces drop out (decrease in memberships) while others increase in memberships and show up. Note the relative sizes of the six clusters. Clusters 1 through 4 are notably smaller than clusters 5 and 6. When physically grouping the plaster face casts into their respective clusters, there is a noticeably larger variance in the forms of the faces in clusters 5 and 6 than clusters 1 through 4. This is due to the nature of the data set. The data set is composed of aircrew members who do not fit into the standard sizing categories. These are the 'fringe elements' of the entire flying population. Clusters 1 through 4 each represent a small population of custom users that are very similar. Cluster 1, for example, is made up of only two members, both of them small females. The faces grouped into clusters 5 and 6 are similar in size within each cluster but still retain variations in form that may be too small to be differentiated by the

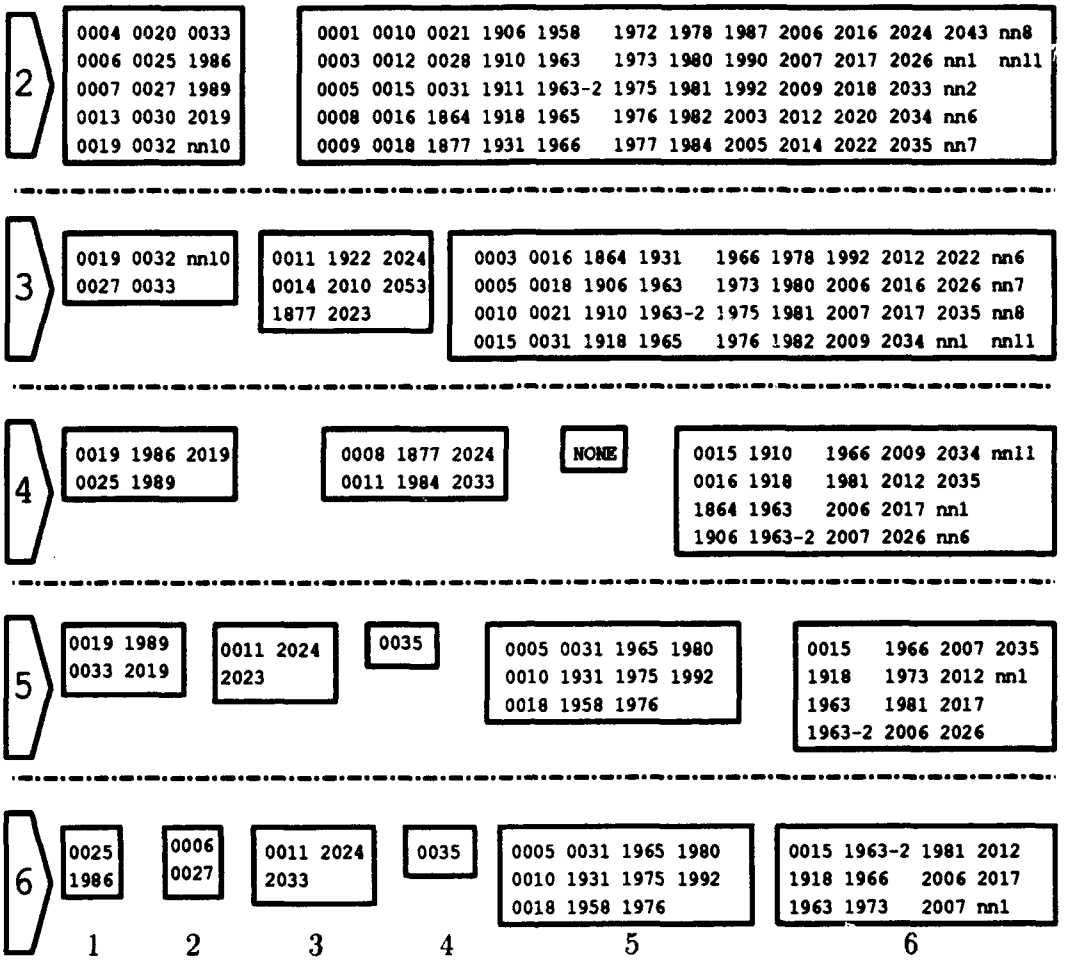


Figure 5.30 Fixed Form Clustering Results: Feature space consists of the top five KLT coefficients from the sparsed, bending energy matrix.

clustering algorithm. This is due to the relative magnitude of the features used by the clustering. It is hypothesized that if the data set is larger, then clustering could be confidently done for a larger number of clusters, and clusters 5 and 6 would split into smaller, more representative clusters.

5.3.2 Face Prototype Selection. To create the six prototype masks, the samples with the highest memberships are chosen from the KLT of the entire sparsed energy matrix to represent the clusters. The six faces chosen to represent these clusters are listed in Table 5.2. Front and side views of these seal areas are shown in

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
1986	0027v	0011v	0035v	1980	1918

Table 5.2 Cluster Representatives: Using results from fixed form clustering, with features from the KLT on the sparsed bending energy matrix.

Figures 5.31 and 5.32. The front views show the outlines of the sparsed seal lines. The points within the seal lines are the 14 spacial delineators. The bridge of the nose is to the right and the chin is to the left. The side views show the seal curves and the 14 spacial delineators as if the wearer is lying face up. The highest point on the right of the side view represents the tip of the nose and the highest point on the left is the chin.

5.4 Prototype Mask Production

One of the objectives of this research is to apply current computer technology in the production of custom and prototype aircrew oxygen masks (AOMs). As detailed earlier in this chapter, face geometry is collected as digitized points, and the data are manipulated to determine cluster prototype faces. This section describes how the digital data representing those prototype faces are utilized in the design and production of hardware that can be worn by operational aircrew members.

Figure 5.33 outlines the process described in this section. The Pro/ENGINEER computer aided design (CAD) system is used to create a CAD model of the plug (see Sec. 5.1.1.1). This model can be based upon the digitally recorded geometry of any selected face. The CAD model is then fabricated using thermoplastic on a computer-controlled rapid prototyping machine (RPM), and the resulting plug is used by the custom mask shop (CMS) to produce latex faceforms for custom MBU-5/P masks.

5.4.1 Computer Aided Design. The following sections explain how computer aided design is used to create a solid image of a MBU-5/P plug. This image is used to create the prototype.

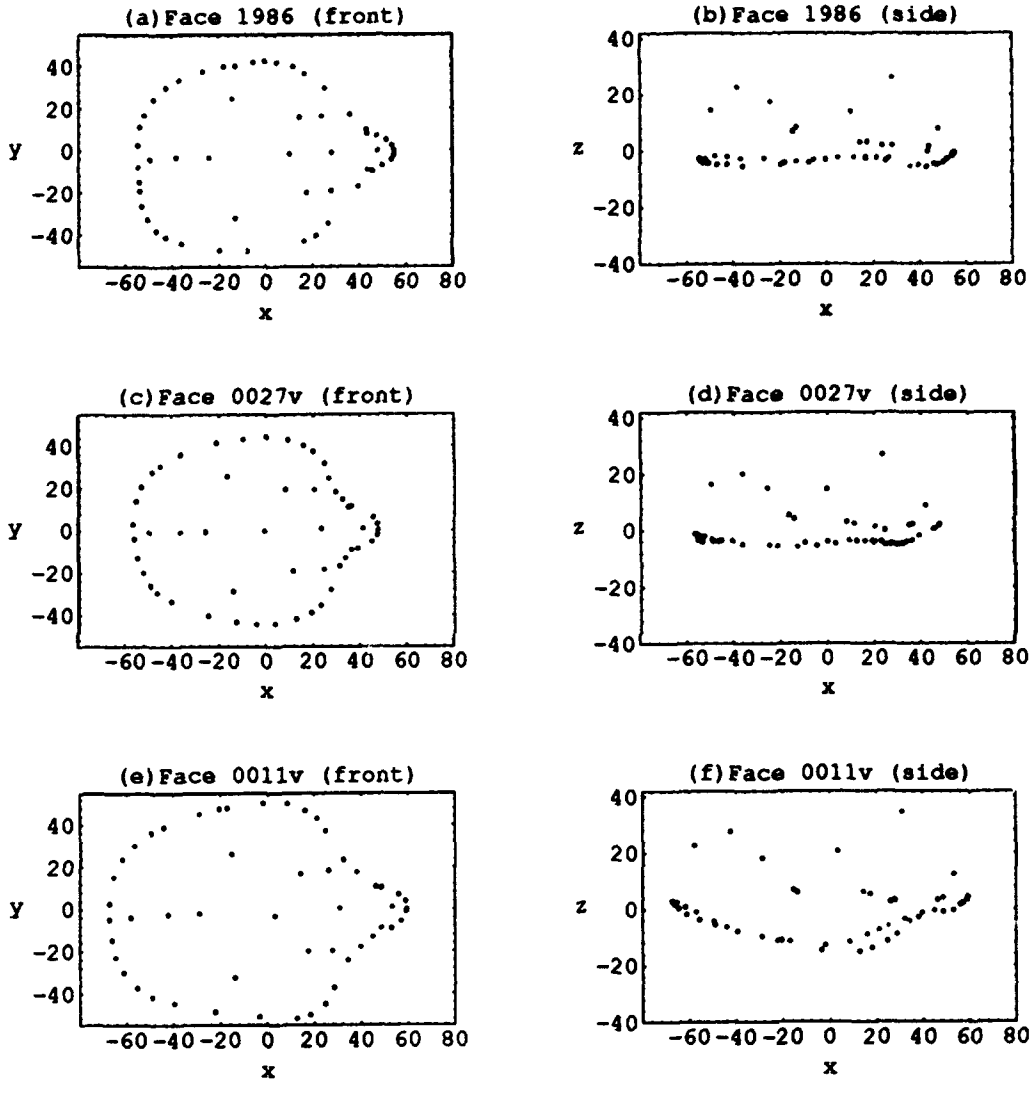


Figure 5.31 Sparsed Prototype Faces: Masks 1986, 0027v, and 0011v.

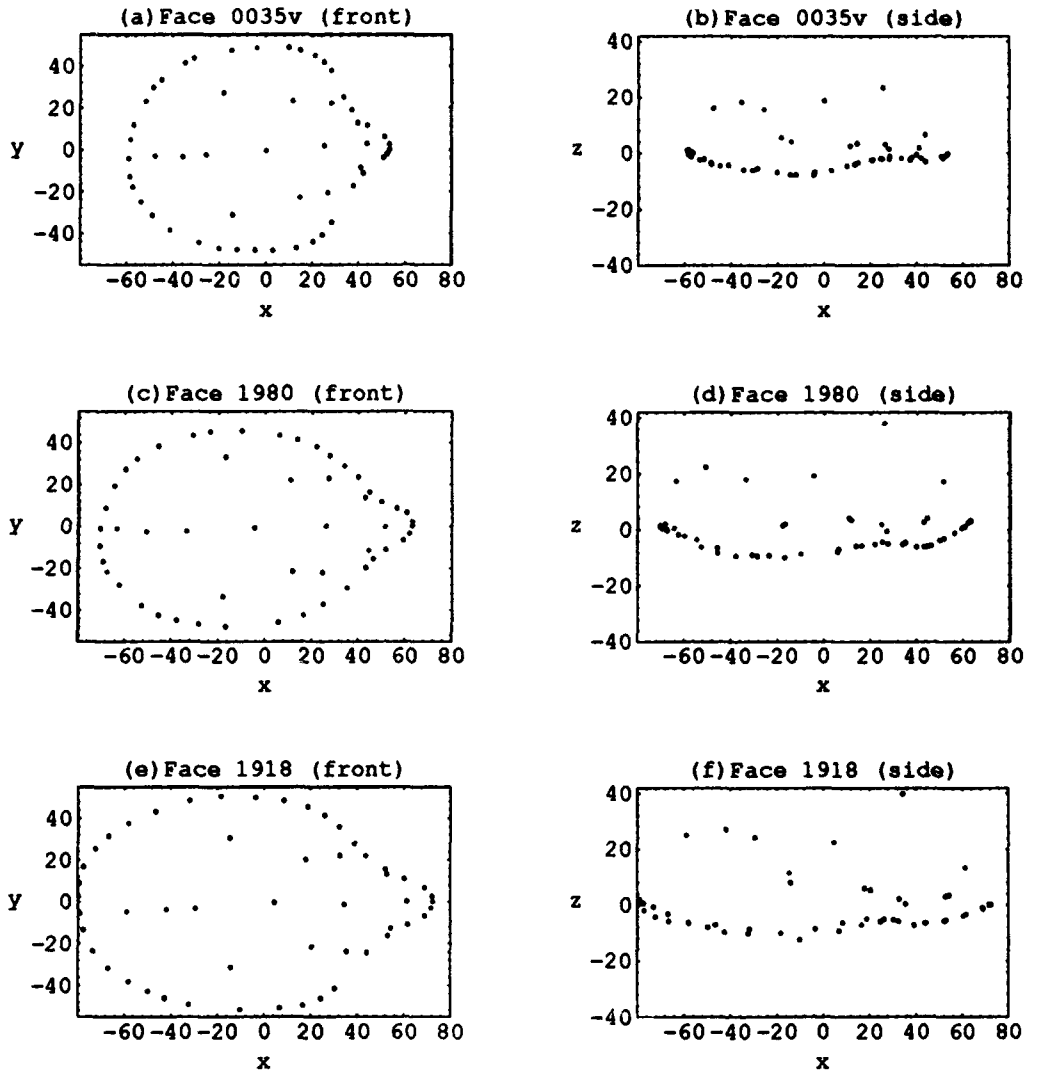


Figure 5.32 Sparsed Prototype Faces: Masks 0035v, 1980, and 1918.

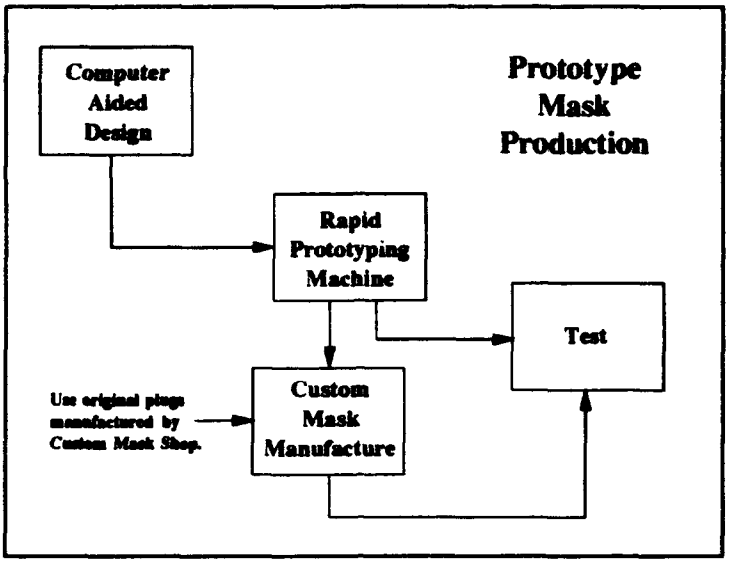


Figure 5.33 Prototype Mask Production Process.

5.4.1.1 Creating the Base Plug Model. The CAD model of the plug is created on the Pro/ENGINEER system, the same system used in the reduction and analysis of digital seal area data (see Sec. 5.1.2). The basic shape of the plug is that of the hard-shell portion of the MBU-5/P mask, so construction of the model begins with reproducing the geometry of the hard-shell in Pro/ENGINEER. Five cross-sections of the hard-shell are digitized in the same manner that the seal lines are digitized from the plaster face casts. These are introduced into Pro/ENGINEER as curves (see Fig. 5.34) and a surface is then created that connects the curves (see Fig. 5.35). Some additional geometry is added to this model so that the plug can accept the metal rod used to hold the plug while it is dipped in latex (52). The result is the 'base plug model' from which all prototype mask plugs are created.

5.4.1.2 Mating the Seal to the Base Plug Model. To create a plug that matches a specific prototype face, the smoothed seal line data file of that face (see Sec. 5.1.2.1) must first be converted into a surface that can be mated to the base plug model. To give the seal line a finite width, an 'outer seal curve' is generated

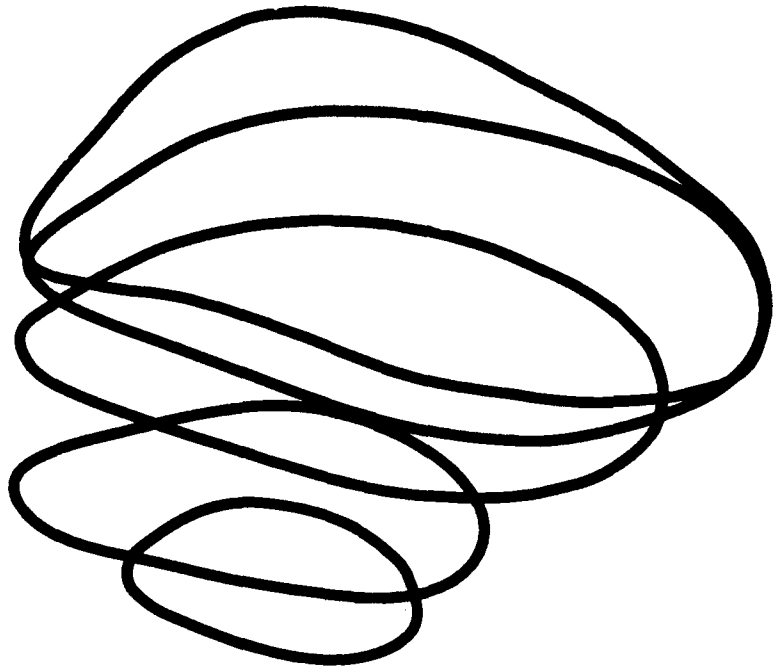


Figure 5.34 Digitized Cross Sections of the Hard Shell.

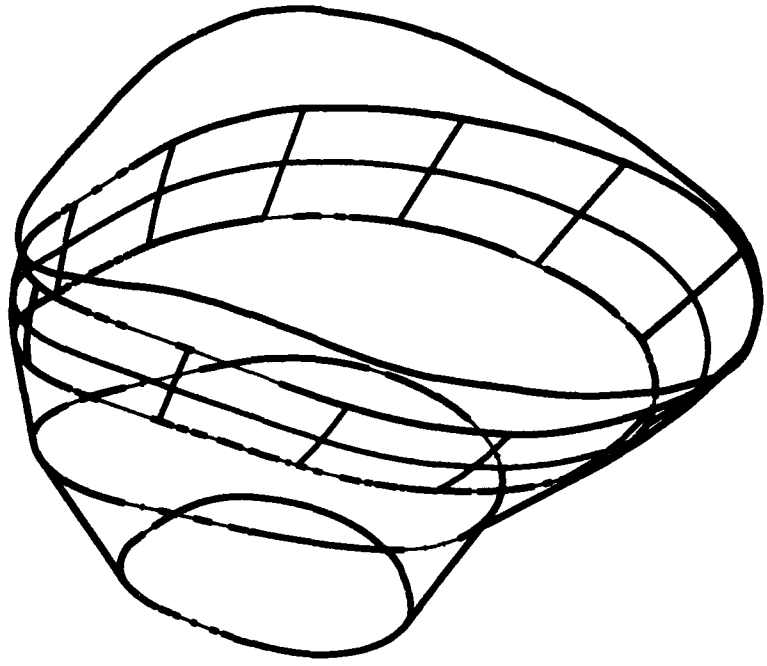


Figure 5.35 Surface Formed from Cross Sections.

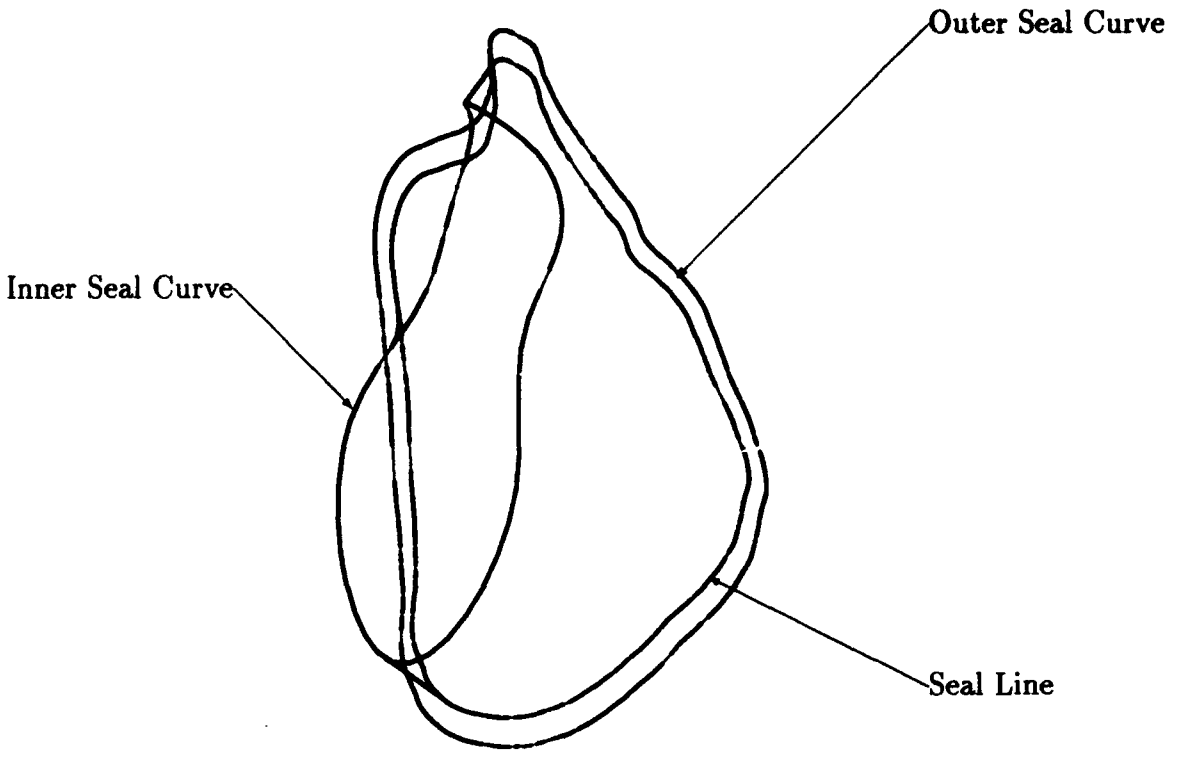


Figure 5.36 Seal Curves.

by offsetting each point in the seal line six millimeters outward from the center. An 'inner seal curve' is also created by drawing a line through ten of the spacial delineators (SDs) in the seal area data file. This curve is an attempt to snugly encircle the nose, mouth, and chin of the face being modeled. Because there is only one SD low on the chin, the inner seal curve is only a rough approximation of the face in this area.

These three curves (the seal line, outer seal curve, and inner seal curve) are connected to form a surface that matches the contours of the prototype face around the seal line and inward toward the nose and mouth (see Fig. 5.36). This surface is aligned with the base plug model, and the two are joined. The result is a CAD model of a plug that is ready to be fabricated. The plug produced will be the functional equivalent of the hydrocal plug currently hand-made by the CMS (as described in Sec. 5.1.1.1).

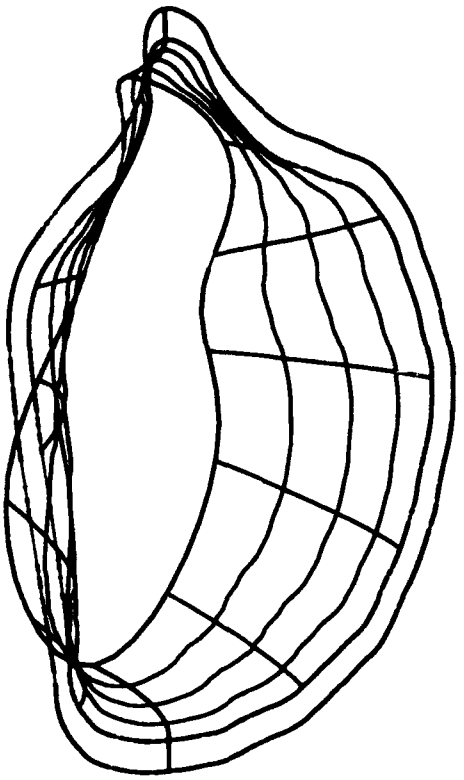


Figure 5.37 Seal Surface.

5.4.2 Rapid Prototyping. The plug model created on the Pro/ENGINEER CAD system is turned into hardware by a rapid prototyping machine (RPM) manufactured by Stratasys, Inc., and known by the trade-name 3D MODELER (51). To prepare the CAD model for fabrication, the RPM system software cuts the model into horizontal 'slices' from 0.05mm to 0.76mm thick, depending on the resolution desired in the model. The slices are formed one at a time by the RPM in a process called Fused Deposition Modeling. Thermoplastic material is extruded from a nozzle that moves in the horizontal plane . The nozzle lays down a 'road' of heated, liquid thermoplastic, ranging from 0.25mm to 3.81mm in width, on a base that moves vertically. On each pass, the nozzle lays down a road that constitutes a horizontal slice of the model. After each pass, the base moves downward a distance equal to the thickness of the slices being produced, so that on the subsequent pass the material is laid on top of the previously formed slice. In this manner, the model is built up

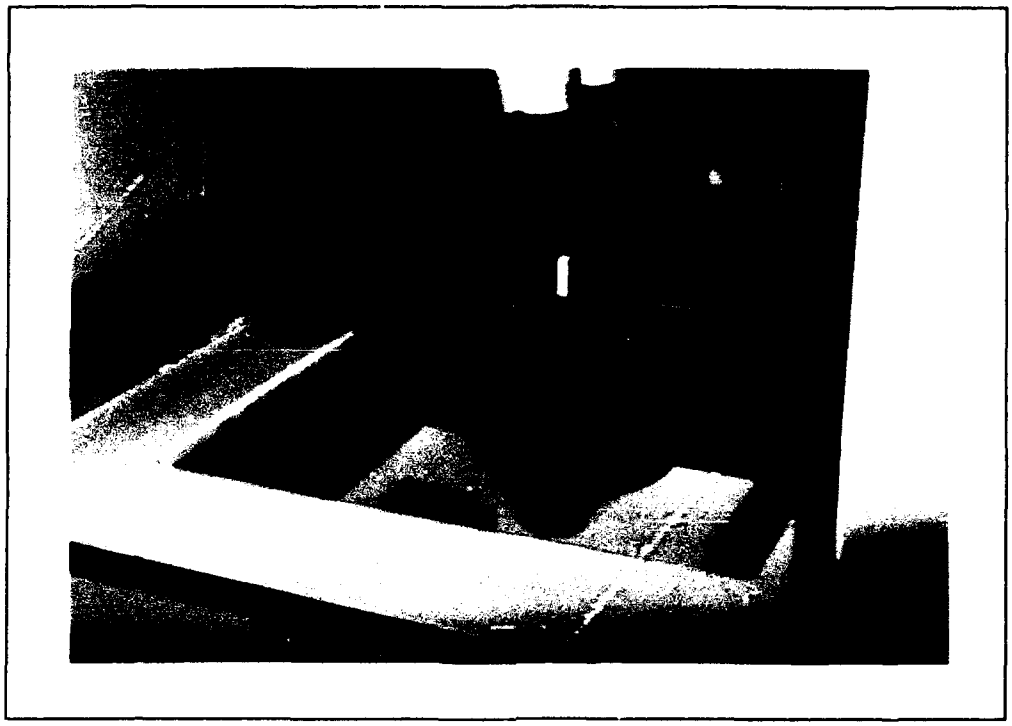


Figure 5.38 The 3D MODELER Rapid Prototyping Machine (RPM).

from the base, with the successive layers solidifying and fusing together as the thermoplastic cools. The overall tolerance of parts produced on the RPM is ±0.13mm, and they have the look and feel of nylon.

To reduce the amount of material and time required to produce a plug, a fabrication mode is used that does not result in a solid block of thermoplastic. Rather, the RPM forms a part with smooth exterior surfaces, but a mesh-like interior. Even with this time-saving mode, each plug requires approximately 72 hours to fabricate on the 3D MODELER.

5.4.3 Prototyping Results. The procedure developed in this research demonstrates the feasibility of using computer aided design and rapid prototyping technologies in the production of custom-made aircrew oxygen masks. The procedure replaces the very labor-intensive process currently used, in which plugs are made by hand from a plaster face cast. The quality of fit achievable with the new procedure

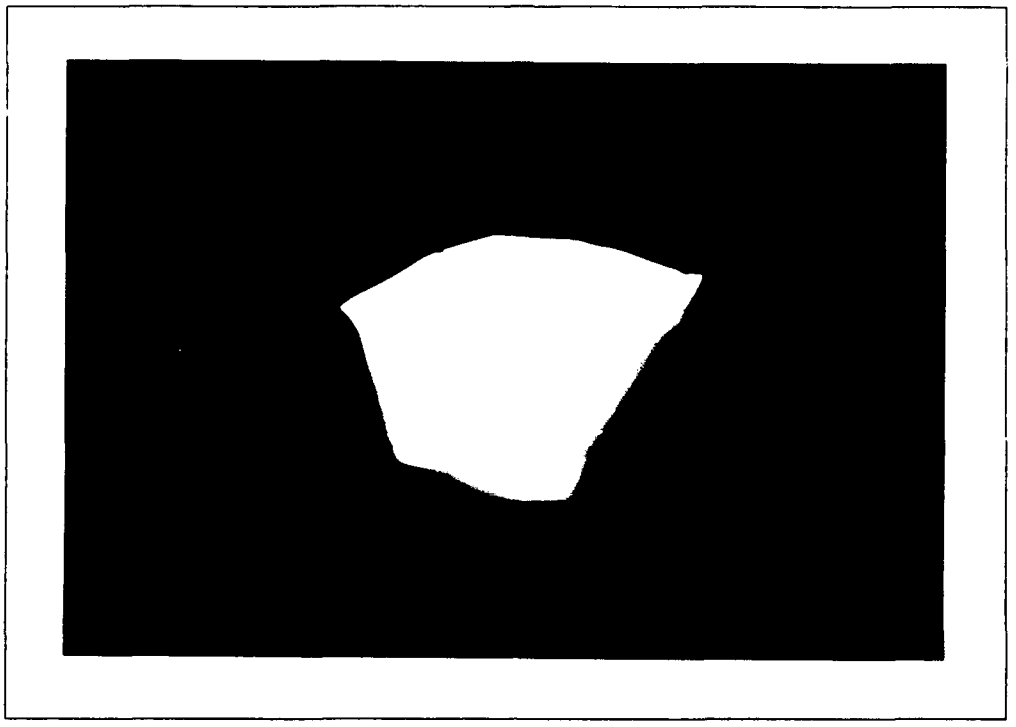


Figure 5.39 Prototype plug fabricated by the RPM.

is assessed to be comparable to the traditional process, although a thorough study of fit is not possible due to time and funding constraints.

5.4.3.1 Time Economy. During the development of this procedure, the production of a single plug required approximately 2 $\frac{1}{2}$ person-hours to create the plug CAD model (excluding form analysis and clustering), and an additional 72 hours to fabricate the plug on the RPM. In a production environment, the 2 $\frac{1}{2}$ hours could be reduced in several ways. First, the manual digitization of face geometry could be replaced by a laser scan. This procedure takes only seconds and has the additional benefit of providing data that may be used to produce a plug that more accurately captures the subject's face (see Sec. 5.4.3.2). Second, an experienced CAD operator can design and implement a custom sequence of commands, streamlining the model-creation phase by reducing the number of operations necessary to mate the base plug model with the seal area. It is likely that these two changes, along with the natural

learning-curve that occurs when a process goes from development to production, would reduce the person-hour requirement to between 1 and 1 $\frac{1}{2}$ hours per plug. This is much less than the 12 person-hours required for the current process.

The limiting factor at this point in the development is the time needed by the RPM to fabricate a plug. The full capabilities of the Stratasys 3D MODELER were not explored during this research, and it is possible that a more time-efficient mode of operation exists. For example, some amount of accuracy in the parts produced could be sacrificed to achieve a reduction in the fabrication time.

5.4.3.2 Quality of Fit. Because of funding constraints, only enough plastic material was available to produce one of the prototype plugs on the RPM. With only one plug on which to make latex faceforms, a rigorous evaluation of how well the faceforms fit their intended wearers was not possible. As an alternative to this, the plug was held against the plaster face cast on which it was modeled, and a visual evaluation made of how well it matched the face cast. This can not, of course, yield any information on the comfort of the mask or the quality of the seal, but does give an idea of whether the procedure results in a plug that is close to the desired shape and size.

The plug appears to closely match the geometry of the plaster face cast around the drawn seal line. This is to be expected, since this seal line is digitized and introduced into Pro/ENGINEER, and forms part of the plug model.

Moving away from the seal line, however, the plug does not match as exactly. In creating the CAD model, only ten data points (selected from the 14 spatial delineators) are used to define the inner seal curve. Recall that this curve is used to form the surface of the plug extending from the seal line, along the cheeks, nose, and chin (see Section 5.4.1.2). This surface is only an approximation of the face in that area, and a particularly poor one in the area around the chin. This is because there are only three spatial delineators below the mouth, much too few to accurately recreate

the geometry of the chin area. To produce a surface that very closely matches the face would require that many more data points defining facial geometry be gathered.

The most practical way to gather the large amount of data required would be with a laser scan of the face. This data could be transferred directly into the CAD program, which would create a surface that exactly duplicates the face scanned. When mated with the base plug model, the resulting plug would produce faceforms that match the face just as precisely as those made from the hydrocal plugs currently used.

5.4.3.3 Faceform Production. The plu· currently used by the CMS are made from hydrocal and sealed with an epoxy finish. The end product of this research is an uncoated plastic plug. To ensure that the plastic plug is compatible with the process used in the production of latex faceforms, a sample plastic plug was sent to the CMS for testing.

Before testing the plug, a material expert at Stratasys, Inc., was contacted and asked if the process would adversely affect the plastic (42). He expressed two major concerns. The first is the ethyl alcohol based coagulant used to coat the plug before the latex dip. Ethyl alcohol is a solvent to many forms of plastic, and could destroy the bonds between the plastic molecules. His second concern was the 60° C (140° F) water bath used to cure the latex is above the 38° C (100° F) glass transition temperature of the plastic. While the plastic would not actually melt, the plug could undergo some changes in shape.

The CMS produced one faceform, and neither the chemical nor the thermal treatment caused any degradation of the plastic plug. However, more testing is required to ensure that the plug will not degrade under repeated use.

Two unexpected problems surfaced in the test production of a faceform. The interior of the plastic plug is formed as an open mesh, and the ethyl alcohol coagulant entered the plug through the mesh. While this did not cause physical damage to the

material, it created a mess and made the plug difficult to store. This problem could be solved by sealing the exposed mesh surfaces with epoxy or a similar material.

The second problem is difficulty encountered in removing the dipping rod. During the faceform manufacturing process the plug is hung by a 6 mm diameter rod while it is dipped in latex and cured in water. After the cured latex faceform is taken off of the plug, the dipping rod is removed for use in another plug. For the test, the dipping rod was glued into a hole in the plug and could not be easily removed afterward. A possible solution is to press a threaded metal insert into the plug and thread the end of the dipping rod so the two can be easily connected and separated.

Testing of the plastic plug shows that it can be used for the production of latex faceforms. There are no major material compatibility problems. The minor problems of sealing the plastic mesh and attaching the dipping rod can be corrected with simple design modifications.

5.5 Evaluation

An evaluation is conducted to show the clustering does indeed work, and to determine a membership cutoff range for the clusters. The evaluation is conducted in two parts. The first is an evaluation of the clusters through use of latex faceforms representative of the cluster centers and the plaster casts of each member in the clusters. The second is a survey in which aircrew members of each cluster evaluate the fit of the custom MBU-5/P mask made for the representative center.

Latex faceforms of the representative centers for clusters 1,3,5 and 6 are made and held to plaster casts in an attempt to evaluate the 'fit'. Clusters two and four cannot be evaluated. The plaster cast with identification number (ID) 0027, the representative center of cluster two, has no known identification; no name; no tracking number; no way to identify the owner of the plaster cast nor any custom MBU-5/P which was made for this person. Cluster four contains only one person,

ID 0035, a person with a very flat face. If a faceform is made for this cluster, there are no other members in the cluster who can be used for evaluation purposes. The faceforms made (clusters 1,3,5,6) are held firmly on plaster casts, and checked to ensure the faceform does not inhibit vision and fits around the chin and nose area. It is noted that the faceform does not 'fit' those with lower membership values, closer to 0.80, as well in the larger clusters, 5 and 6. Deviations of fit occur around the chin area and obstruct the view in some cases. Although a guess can be made about fit in this manner, no information can be gained as to where leakage may occur during positive pressure air breathing. Holding a latex faceform to a plaster face cast is much different than strapping on an entire MBU-5/P mask tight enough to preclude most air leakages during positive pressure breathing. The preferable way to evaluate fit of a mask is having an aircrew member breath positive pressure air to test for air leakage. To accomplish this a survey was conducted.

5.5.1 Survey. A complete custom made MBU-5/P mask is made for the representative centers of clusters 1,3,5 and 6. The representative centers have the highest membership values in each cluster. The custom masks made for each cluster are sent to the person with the next highest membership. Table 5.3 shows the identification numbers and membership values for the representative centers and the survey personnel. Appendix E contains the representative center and survey personnel data plots for each cluster, and photographs of each representative cluster center plaster face cast. Survey personnel are asked to take the custom mask to their life support office, which fits all aircrew personnel with masks, and test the mask for air leakages by breathing positive pressure air. Questionnaires accompanying the mask are completed by the survey subjects and their answers are provided in Appendix D.

All four test masks have been evaluated by the survey subjects breathing positive pressure air through the custom MBU- 5/P mask representative of the cluster centers. The survey results are split: one test mask fit well and the other three did

not. Two survey subjects whose test masks did not fit well are from the larger clusters 5 and 6. These subjects comment that the masks are too large, obstruct their view under positive pressure breathing, and leak around the eye/nose area. Cluster 3 test subject comments that he had difficultly fitting the mask around the nose area. The test subject from the smaller cluster 1 comments that the test mask is suitable and requests the test mask as a replacement to her own custom MBU-5/P mask.

Although an extensive survey has not been conducted, the results suggest that the membership cutoff values for fit within each cluster may be affected by the inherent variability of form between the members within each cluster. The survey subjects from the larger clusters 5 and 6 have memberships of 0.88 and 0.93 respectively, yet the subjects did not find the test mask adequate. This suggests that the membership required for fit is larger than 0.88 for cluster 5 and 0.93 for cluster 6. The larger variability in form within these two clusters could be due to the uniqueness of the entire data set. The clustering algorithm inherently groups similar seal areas within a cluster and identifies different clusters which have large dissimilarities between clusters. Clusters 1,2,3, and 4 contain very few seal areas, while Clusters 5 and 6 are much larger. If the clustering algorithm separated the most unique seal areas into clusters 1-4, that leaves the majority of the seal areas in clusters 5 and 6 which cannot be reliably broken into smaller clusters due to the small sample size.

It is important to note a few facts concerning cluster 3 survey subject. First, the subject found the overall size of the test mask is good, but leaks occur around the bridge of the nose under the eyes while breathing positive pressure air. The mask can be tightened so the leaks stop, but the subject experiences pain in the nose area where the subject has broken his nose several times. Secondly, the plug used to create the faceform of the test mask did not come from the representative center of cluster 3. The representative center of cluster 3 is a Category V, so there

was no CMS plug created from this plaster face cast. The CMS did not have time to create a new plug from ID 0011 plaster cast (the representative center) in order to create the test mask faceform. The plaster face cast which was used to create the plug utilized in manufacturing the test mask faceform has long ago been destroyed; it was not available to digitize and add to the analysis to determine whether or not it too would have been classified within cluster 3.

The survey subject from cluster 1 having a membership of 0.84 found the test mask to fit well. This implies the representative center mask would fit other faces in the cluster with memberships above 0.84. This survey subject was classified as a Category V when her plaster cast was received by the CMS. This means another plug was used to create the faceform of her custom MBU-5/P mask. This indicates that the form characterization method developed in this research was able to determine a better Category V plug than the CMS.

5.5.2 Survey Summary. The survey demonstrates that membership cutoff values for fit within each cluster may be affected by the inherent variability of form between the members within each cluster; evidenced by the shape of the clusters. Due to the small sample size, breaking clusters 5 and 6 into a larger number of clusters is not possible. Time constraints of the research and CMS precludes a faceform from the representative center of cluster 3 (ID 0011) from being manufactured; thus cluster 3 has not been adequately tested. Testing of cluster 1 shows that the method presented in this research clusters seal areas better than the CMS can by visual inspection, and the trial and error method of fitting existing plugs to plaster face casts.

Cluster Number	Rep Center ID Number	Membership	Survey Personnel ID Number	Membership
1	1986	0.92	0025	0.84
2	0027	0.96	–	–
3	0011	0.94	2024	0.89
4	0035	0.91	–	–
5	1980	0.93	1958	0.88
6	1918	0.99	1981	0.93

Table 5.3 Representative Centers and Survey Personnel Cluster Memberships

VI. Conclusions and Recommendations

This chapter consists of conclusions made in regards to this research, and recommendations both for continued research using the MBU-5/P data (project recommendations) and for future studies using the energy method developed.

6.1 Conclusions

The major contributions of this project were the ability to characterize form through the use of spacial correlation in the form of the energy matrix, the use of fuzzy clustering to classify the human face into groups for aircrew oxygen mask designs, and the creation of a prototype plug from the information gathered on the seal areas. The end product was an ability to create face masks from the cluster representatives which better fit the population of custom mask users. All this was done without prior knowledge of the anthropometrics or demographic information of the aircrew members: pre-segregation was not possible since the source of data consisted of plaster face casts, some without identification regarding sex, and with no additional information regarding age or size. The input data consisted only of three dimensional data points, which were chosen based on the best representation of the seal area under investigation.

The energy matrix, based on a linear system of equations using the variogram function, enabled the prediction of values in areas where there were no measurements. This was the source of spacial correlation and surface prediction which provided features for clustering. The difficulty lay in data reduction and identifying the critical features for comparison. The Karhounen-Loeve Transform performed reasonably well and provided features of greatest variance for cluster determination. Fuzzy clustering enabled the selection of the cluster representatives and tariffs based on membership values. Computer aided design combined with rapid prototyping technologies allowed for the recreation of the seal area. The final proof of the ap-

plicability of the entire process to grouping custom MBU-5/P mask users involved the contributed efforts of the custom mask shop to send masks based on the cluster representatives to other members within the cluster. The general response is that the ability to identify similarities between individuals is reasonably high for small clusters, but goodness of fit decreases as variability of form between members within each cluster increased.

The contribution to the custom mask shop, is that CMS could use the cluster prototypes as a source of Category V masks. The Category V masks should then fit better, providing a better quality product to the aircrew member. The result is a potential reduction in the number of plugs that have to be produced. This research addressed the problem of the current aircrew face masks failure to fit the increasingly diverse population of aircrew members, by providing a new sizing method, which in turn would reduce the efforts of the custom mask shop. The research was organized by incremental goals or objectives to develop a sizing method sensitive to the needs of the Air Force. The following paragraphs address conclusions developed as a result in attaining the research objectives.

Collect and Prepare Data. The first goal was to collect as many samples as possible of the only sample data available to this research, the custom MBU-5/P mask users: aircrew members who do not fit standard sizes. This research utilized plaster casts made for custom MBU-5/P mask requests and employed a state-of-the-art, three-dimensional digitizer to collect seal area data for use in form characterization analysis. This seal area data were pre-processed by smoothing the seal line and orienting it for computational convenience and visual ease in CAD programs. Orientation of the seal area data eased structure analysis, but was not required for utilization in the form characterization analysis developed in this research.

Develop Form Characterization. It was necessary to develop a form characterization method which captured all the critical features necessary for comparison, was coordinate free, independent of landmarks, did not require a reference form, and could reconstruct the form of the object under study in order to reach the next goal of this research. Multivariate statistical methods used in the 1950's Air Force form analyses for air-breathing oxygen masks could not consistently reconstruct the face from the linear dimensions collected. The linear dimensions were measured between landmarks thought to distinguish different characteristics of the face. There were several other morphological methods available for form characterization but were not applicable to this research because they were landmark dependent, did not provide an accurate surface description between the defined landmarks, and required a reference or target face to compare multiple faces. These were: procrustes, finite element scaling analysis, Euclidean distance matrix analysis, and thin plate splines. Kriging was the only method discussed in this research which included the spacial correlation of the data, is not landmark dependent, and does not require a reference form.

The form characterization method utilized spacial correlation as determined in the structure analysis portion of kriging. Kriging combined spacial correlation with linear equations for optimal prediction while minimizing estimation error. The foundation for providing the spacial description was the variogram function which depends on the linear distance between locations. The result was an energy matrix which captures the form of a face without the use of landmarks.

The energy matrix uniquely characterized the form of a face and provided an large+ number of possible feature sets to cluster. The 'curse of dimensionality' permitted only a few dimensions, or features, per seal area be utilized in clustering. Two methods of data reduction, free form and fixed form, were explored to find the feature set which yielded the best results. The advantage of using free form methods was that seal areas could contain an unequal number of data points. The

disadvantages were the existence of a large data set, coupled with the difficulty of identifying the critical features for comparison. The fixed form method allowed the application of linear transformations, such as the KLT, which could be applied on the entire energy matrix of the sparsed seal areas, the eigenvalues of the energy matrix of the sparsed seal areas, or to a selected uniform number of eigenvalues from the unsparsed, free form data set. The KLT yielded sets of ordered, orthogonal, uncorrelated features. Many different feature sets were explored in the fuzzy c-means algorithm used for clustering, but the fixed form KLT performed on the entire sparsed energy matrix provided the best clustering results. It was postulated that more information was lost during data reduction in taking the KLT on the eigenvalues of the energy matrix than in performing the KLT directly on the fixed form energy matrices.

Clustering. Six clusters were produced from the 99 sample seal areas using only five features: the top five KLT coefficients from the fixed form energy matrices. The seal area having the highest membership within each cluster became the representative for the corresponding cluster. The representative seal area data were fitted to a hard-shell in a computer aided design (CAD) program and a three-dimensional computer representation of a complete plug was developed. Surface information was then input to a rapid prototyping machine capable of producing prototype plugs made of polyamide or polyolefin. Due to funding constraints, the amount of prototyping material available was only enough to produce one prototype plug. The procedure developed in this research demonstrated the feasibility of using CAD and rapid prototyping technologies in the production of custom-made aircrew oxygen masks.

Evaluation. An evaluation was conducted to demonstrate that the energy method developed, the feature set chosen, and resulting clusters did indeed meet the goals of this research. The evaluation was executed in two parts.

First, latex faceforms of the cluster representatives were fitted to the plaster face casts of each member within their cluster. Secondly, custom MBU-5/P masks made from the cluster representatives were sent to the aircrew members in their cluster for fit testing. The aircrew members were asked to test the masks for comfort and air tight seal under positive pressure breathing conditions. The evaluation showed that the cluster membership value required for fit could have been affected by the inherent variability of form between members within each cluster.

The entire method, from form characterization to prototyping, presented in this research achieved the objectives set forth, but was limited in the extent of its success. The seal area data used were not landmark dependent, but the same number of data points per seal area were required in order to perform the KLT on the entire energy matrix. Other limitations were due to the number of sample seal areas available, the sample space (custom MBU-5/P users) itself, and the number and placement of spacial delineators. These limitations are explored further as project recommendations in the following section.

6.2 Project Recommendations

A limiting factor on this research is the uniqueness of the sample space. The custom made MBU-5/P masks are for the unique aircrew members who cannot be fitted with standard masks. Using this 'fringe' data leads the research to attempt clustering on individuals who should be 'unclusterable' by the standard fitting MBU-5/P. Despite this fact overshadowing the research, there are alternatives to evaluate, continuing research in this area by: collecting more plaster casts as they become available; using more spacial delineators; applying different data reduction methods; choosing different feature space selections; and sealing the prototype plug before faceform production.

Sample Size. As the Custom Mask Shop (CMS) receives more masks, the sample size increases. The dimensionality constraints would still apply but the increased sample size allows for a larger number of features and/or larger number of clusters to be used in the analysis. This provides for the classification of more faces together and the possibility of more prototype plugs which fit a larger population.

Spacial Delineators. This research could be continued by adding more spacial delineators within the seal areas, particularly in the area of the chin. Adding more data points will enable a better description of the bending energy required to deform a plane through the seal area data, which better captures the form of the face. This results in more descriptive energy matrices and thus a better feature space for clustering.

Data Reduction Method. The energy matrix contains all the information necessary to describe the form of a face. This research investigates the spectral decomposition of the bending energy matrix, the eigenvalues of the bending energy matrix and sparsing of the original seal line data as methods to reduce the data. In this research the entire energy matrix was used during the analysis. However, taking the inverse of $\mathbf{L}$ is a computationally expensive process. The process of finding features could be made much more computationally efficient if equivalent features were extracted directly from $\mathbf{L}$ without taking the inverse. A partial simplification would be to take advantage of the fact that the energy matrix is symmetric. For a symmetric matrix, Cholesky factorization gives $\mathbf{A} = \mathbf{R}^T\mathbf{R}$ where $\mathbf{R}$ is an upper triangular matrix (50). The inverse of a triangular matrix requires significantly less steps then a 'full' matrix. These are not the only data reduction methods available. Others could be tried (43, 47, 55).

Feature Sets. Along with different data reduction methods, different feature sets could be attempted. Many different feature spaces are investigated in this research, but only the KLT of the bending energy matrix gave accurate results. The KLT is used to identify the most unique features for a fixed form data set, simplifying the selection of which features to use. Other transformations, such as Fourier series, Eigensystem, and wavelets are available.

Prototype Plug Modifications. The procedure developed in this research demonstrates the feasibility of using computer aided design and rapid prototyping technologies in the production of custom-made aircrew oxygen masks. The process significantly reduces the time required to produce a custom MBU-5/P mask. The prototype plug fabricated on the Stratasys rapid prototyping machine (RPM) accurately reproduces the geometry of the face from which it is modeled, but only in those areas where sufficient digital data is available. For the RPM-produced plug to be as accurate as the hand made plug it attempts to replace, a more complete digital representation of the face is needed. This digital data could come from a laser scan of the face. The prototype plug tested shows that the material used to form the plug withstands the chemical and thermal treatments involved in producing latex faceforms. Only one modification, that of sealing the mesh within the plug, is needed to keep the chemicals from seeping into the interior of the plug.

6.3 Recommendations for Future Research

Future research is recommended in applying the methods developed in this thesis to sample data sets other than the custom MBU-5/P users, such as COMBAT EDGE helmets or other aircrew equipment, and to further investigate the methods developed here.

Sample Data Set. A recommended sample data set is a 'normal' data set with as many objects as possible. A 'normal' data set includes a homoge-

neous representation of the objects being analyzed. For further research in the area of air-breathing oxygen masks, the methods developed in this research should be applied to the whole population of aircrew members to better evaluate the information carried in $\mathbf{L}_n^{-1}$ and to develop a better sizing system for the Air Force. This new sizing system could be used to determine better designs for air-breathing oxygen masks currently under development such as the MBU-20/P, COMBAT EDGE.

Further Investigation of Free Form Analysis. The free form method is advantageous because it allows an unequal number of data points per object, but its disadvantages are the large data set and the difficulty of determining which features best characterize the object form. The validation methods using the 'eggs' and the verification by reconstruction of the seal areas using bending energy modes both justify using 'all' the eigenvalues of the bending energy matrices. However, the 'curse of dimensionality' permits only a few dimensions to be used in the clustering routines. Due to the nature of the seal area data, a sparsing routine enabled the reduction in data to a uniform number. Given a different type of data collection method, such as laser scans, this may not be possible. Therefore, a method for evaluating free form derived features in order to identify which are critical is necessary.

Adapting Method to Use Laser Scan Data. The technology exists to laser scan entire parts of the human body, such as the head. Laser scans provide much more data than that obtained from digitizing the seal areas from plaster casts. Future research could be directed toward developing a method to use laser scan data as input into the bending energy method. The method would need to address minimizing computational time and memory, and data reduction methods already discussed would be required. Most of these techniques have been developed in previous studies and could easily be incorporated in future efforts.

Project recommendations and future recommendations for further research involve adapting the energy matrix to different types of data bases, such as laser scans. The ensuing problem is the large matrix which magnifies the importance of data reduction and feature space reduction. However, the form characterization method is applicable to any size data base, and is fairly simple to use. This research provides the basis for future aircrew oxygen mask sizing and designs which would better fit the larger population of aircrew members.

Appendix A. Bending Modes

This appendix contains the complete set of bending modes for mask identification number 0002. The derivation of bending modes is given in Section 3.4.2. The derivation shows that the sum of the bending modes is equivalent to the bending of a flat plate to the face. The sum of the modes plus the regression plane which they are bent from is the best linear unbiased predictor of the face.

Figures A.1 through A.50 are the fifty bending modes for mask identification number 0002. Figure A.51 is the sum of the fifty modes. Figure A.52 is the regression plane which all the modes are bent from. Figure A.53 is the sum of the modes plus the regression.

For the facial data considered in this research, the BLUP is based on the residuals of the face. The residuals are created by subtracting the second order trend from the facial coordinates. Adding the trend to the residuals gives the recreation of the face.

The polynomial trend is given in Figure A.54. Figure A.55 is the addition of the polynomial surface and the residuals. This is the estimated height of the object in physical dimensions.

The face is oriented with the nose pointing 'down' in the $-z$ direction. The approximate location of the bridge and tip of the nose are $x = 50$, $y = 0$, $z = -30$, and $x = 30$, $y = 0$, $z = -30$, respectively.

Figures A.56 through A.65 show a slice of the face through the plane $y = 0$. The same summation process is shown in these figures as in the surface plots. The result is a profile of the nose and chin.

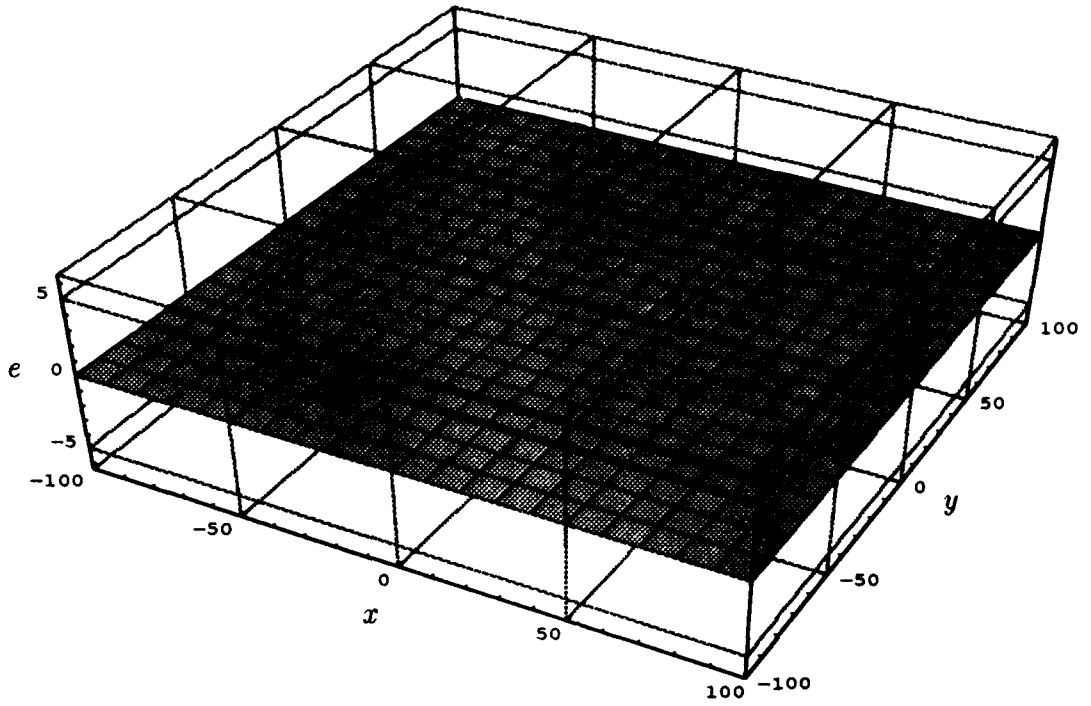


Figure A.1 Bending mode for $\lambda_1 = -0.73799$.

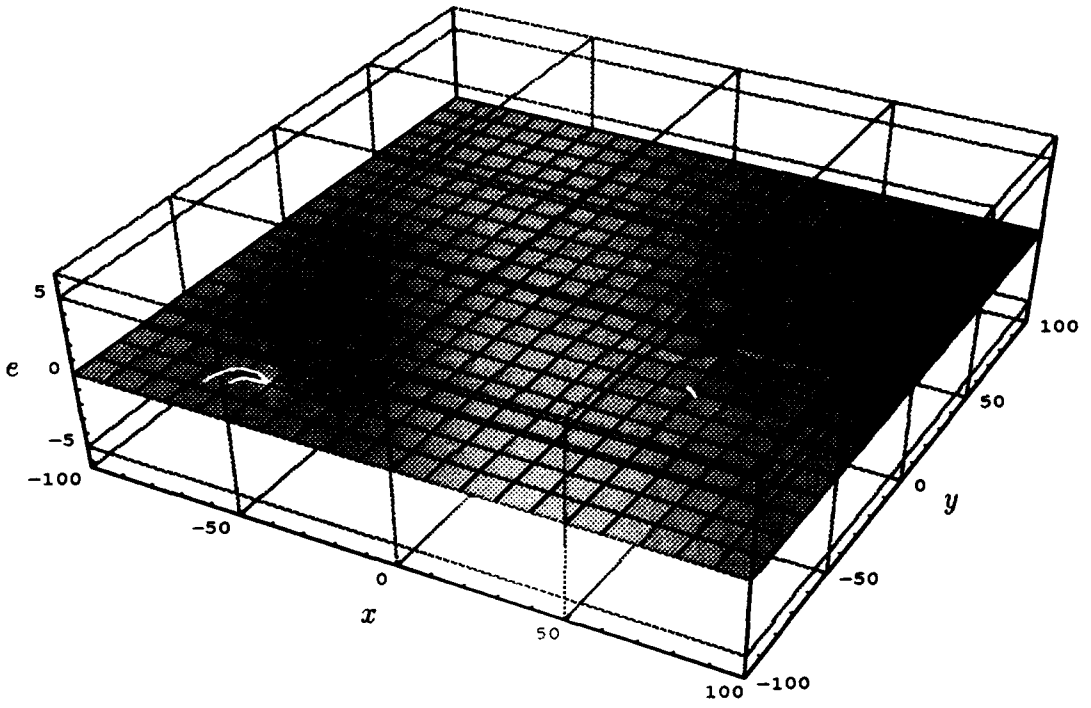


Figure A.2 Bending mode for $\lambda_2 = -0.66514$.

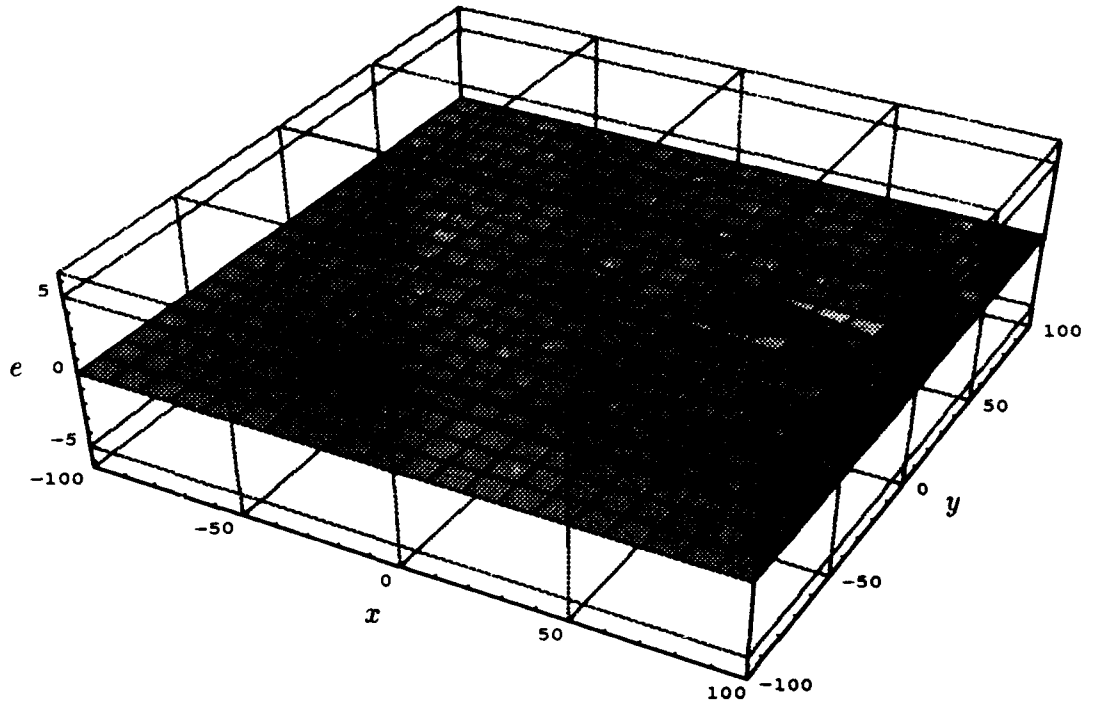


Figure A.3 Bending mode for $\lambda_3 = -0.46310$.

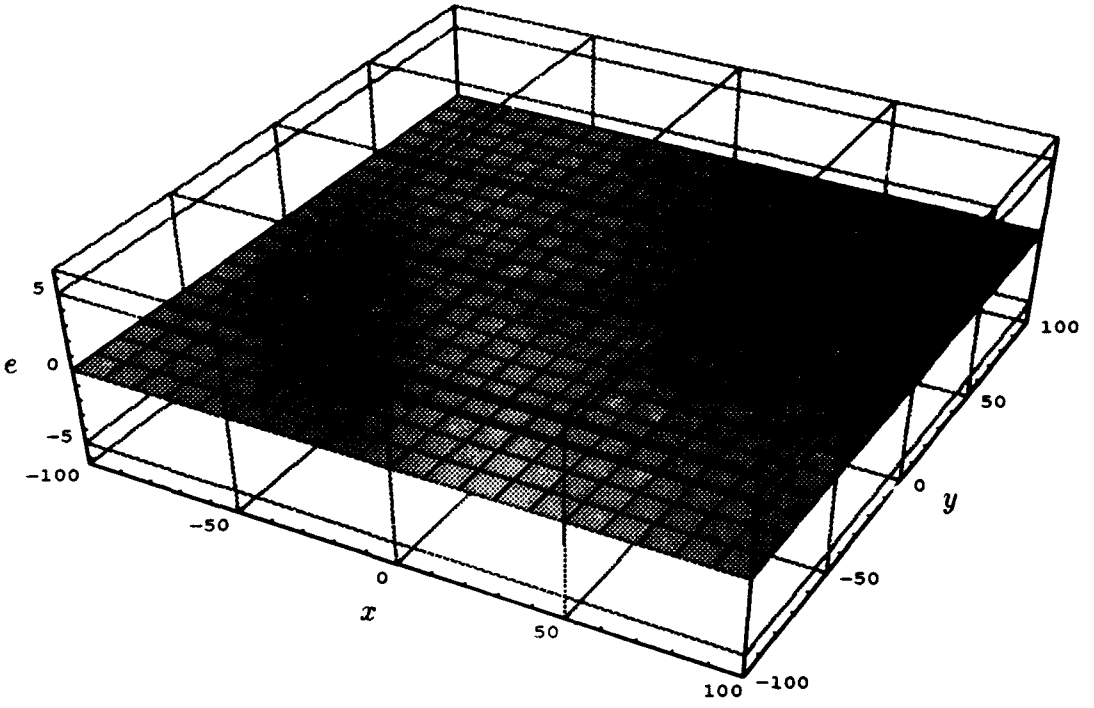


Figure A.4 Bending mode for $\lambda_4 = -0.45198$.

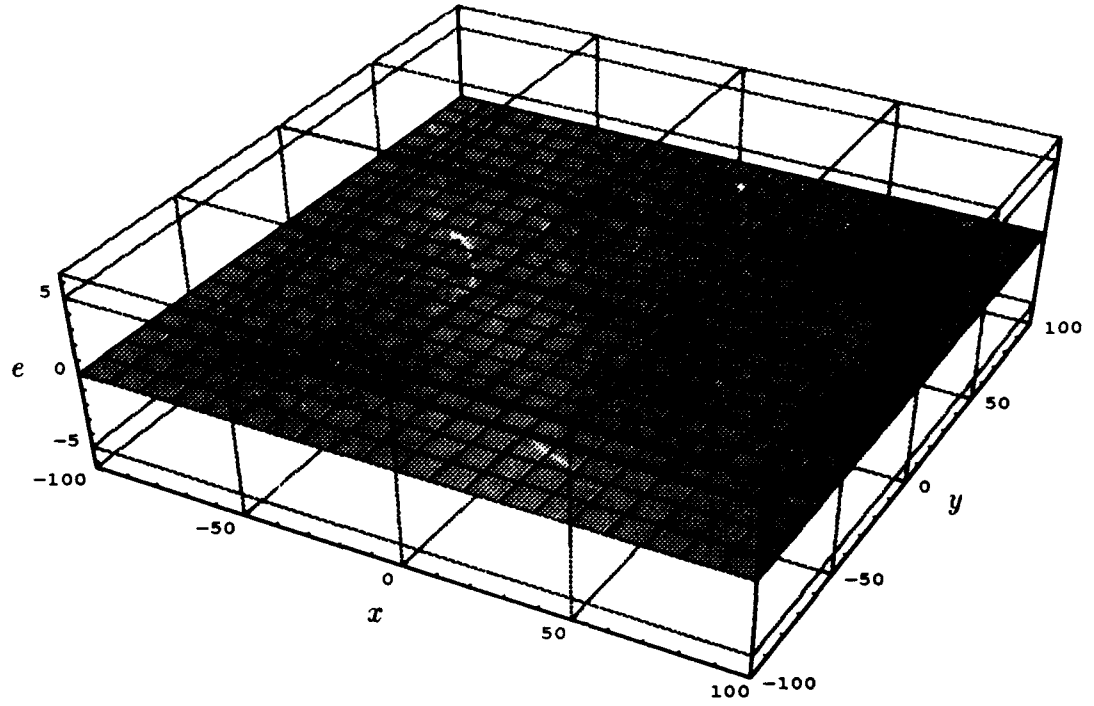


Figure A.5 Bending mode for $\lambda_5 = -0.36815$.

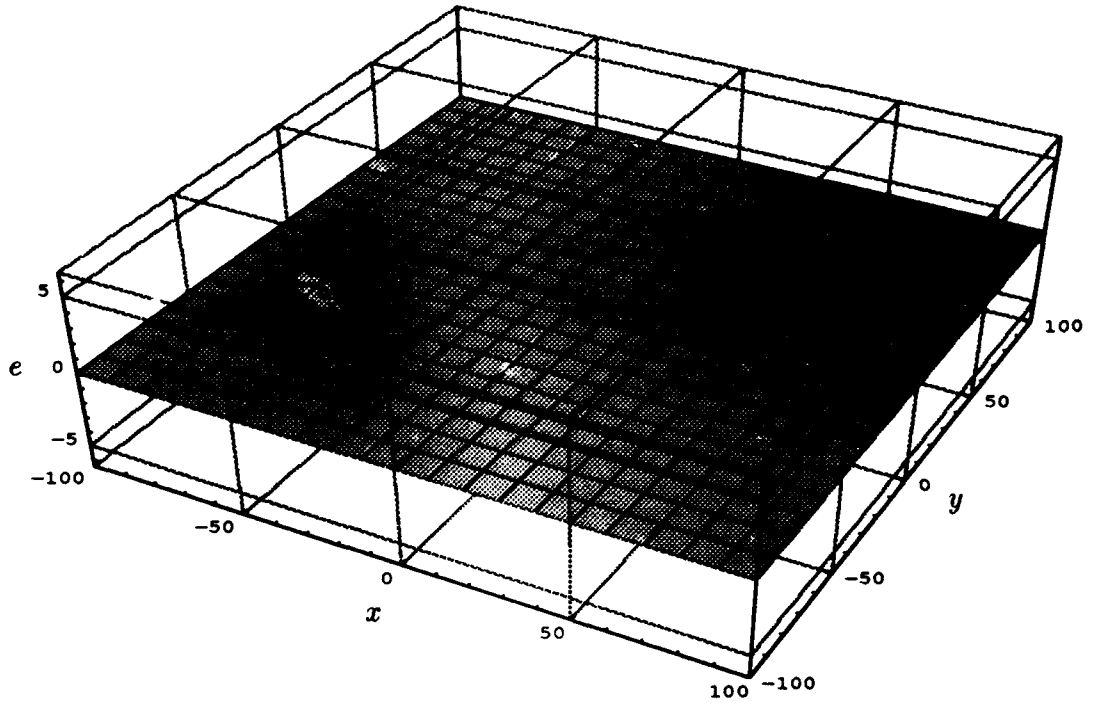


Figure A.6 Bending mode for $\lambda_6 = -0.34010$.

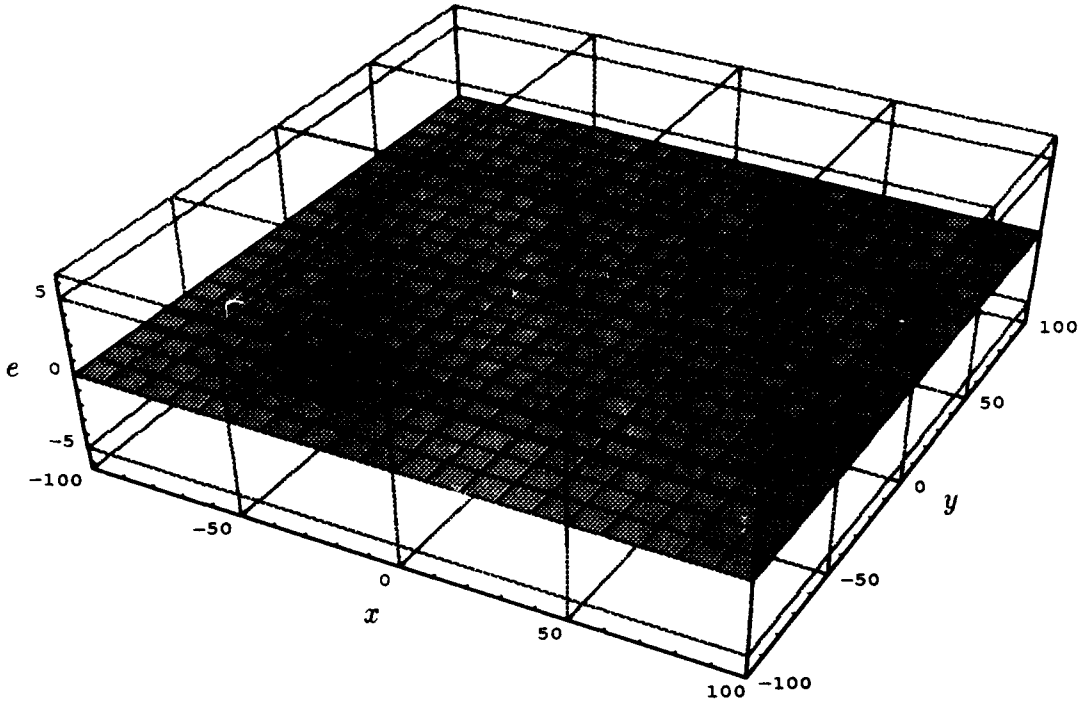


Figure A.7 Bending mode for $\lambda_7 = -0.33584$.

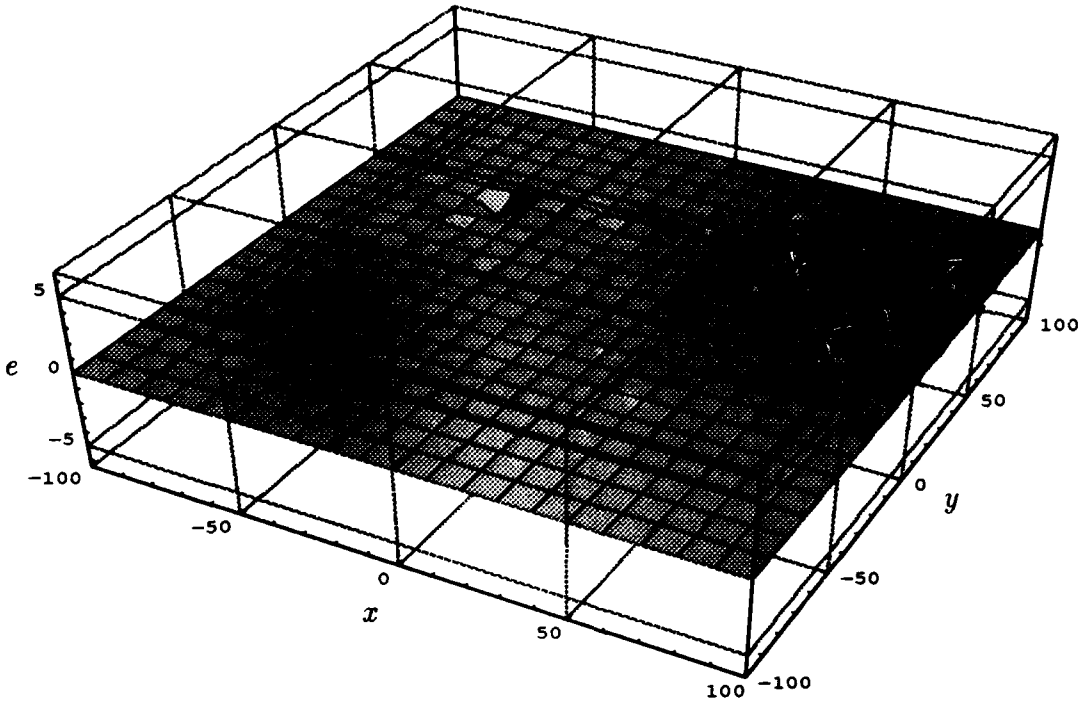


Figure A.8 Bending mode for $\lambda_8 = -0.29655$.

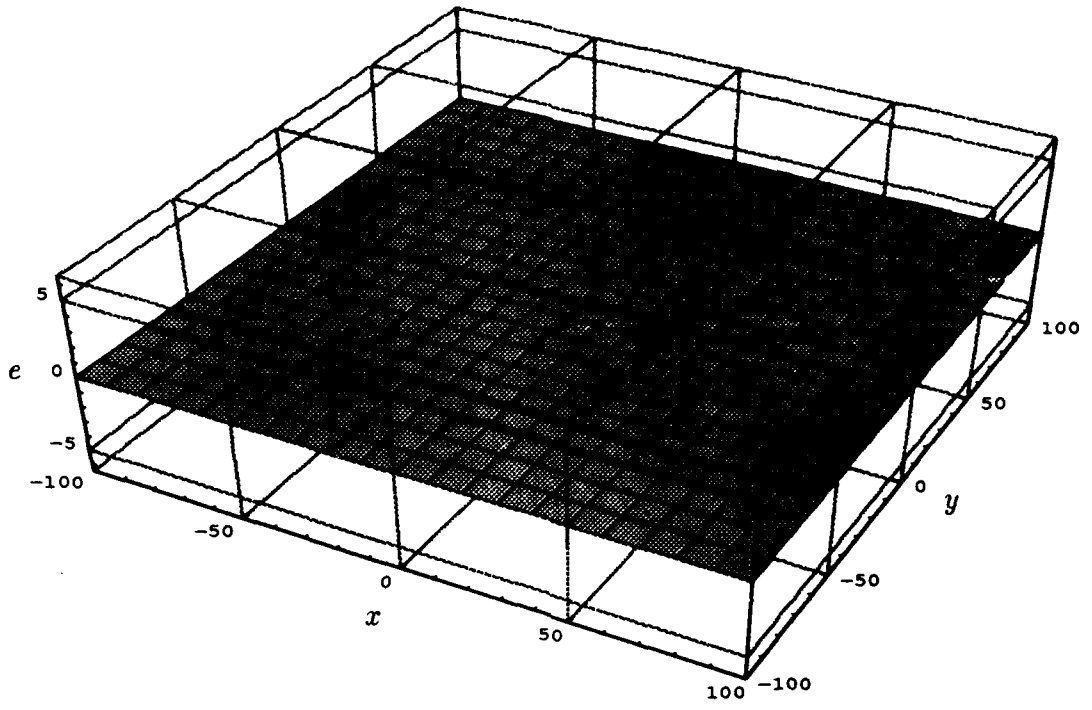


Figure A.9 Bending mode for $\lambda_9 = -0.29053$.

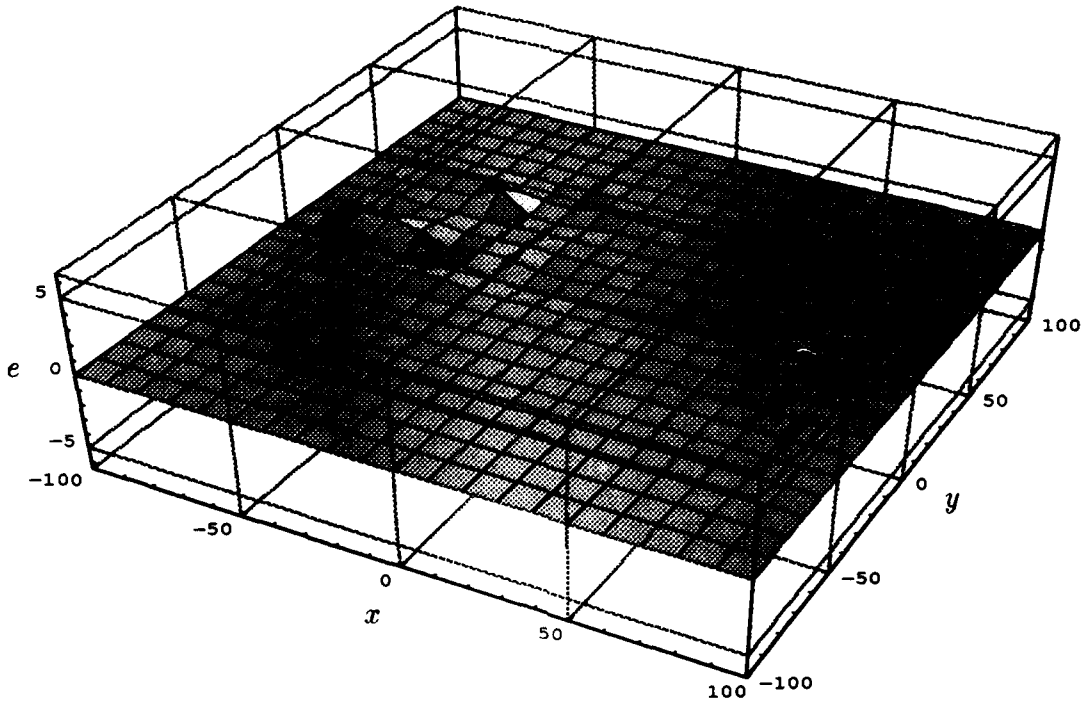


Figure A.10 Bending mode for $\lambda_{10} = -0.28155$.

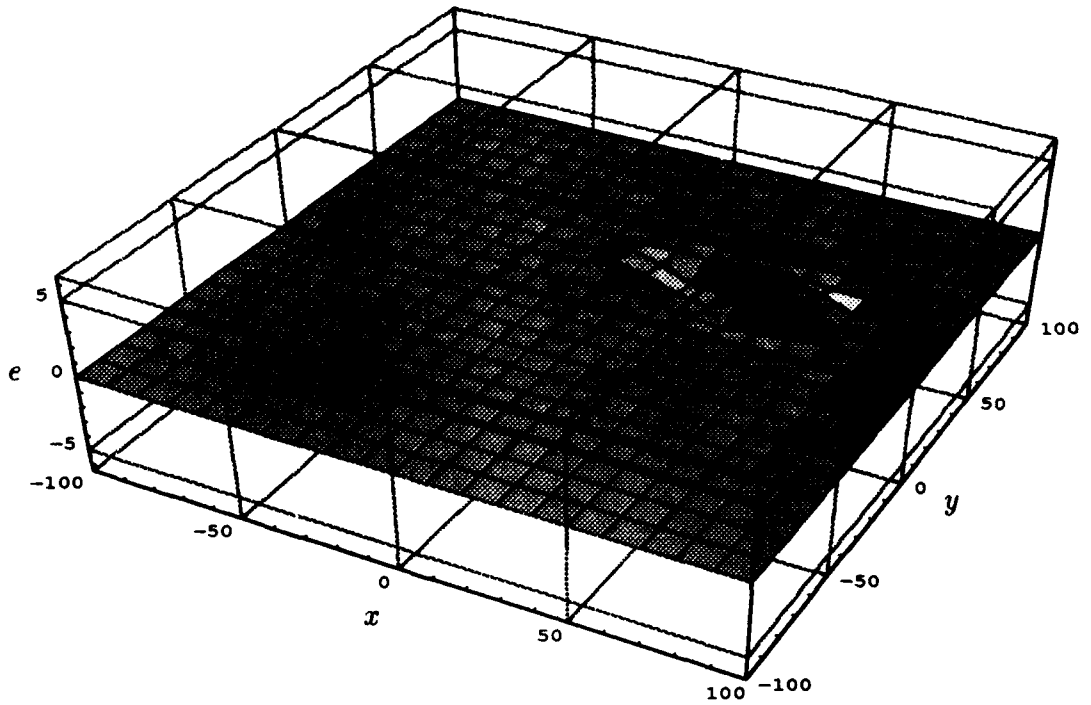


Figure A.11 Bending mode for $\lambda_{11} = -0.26852$.

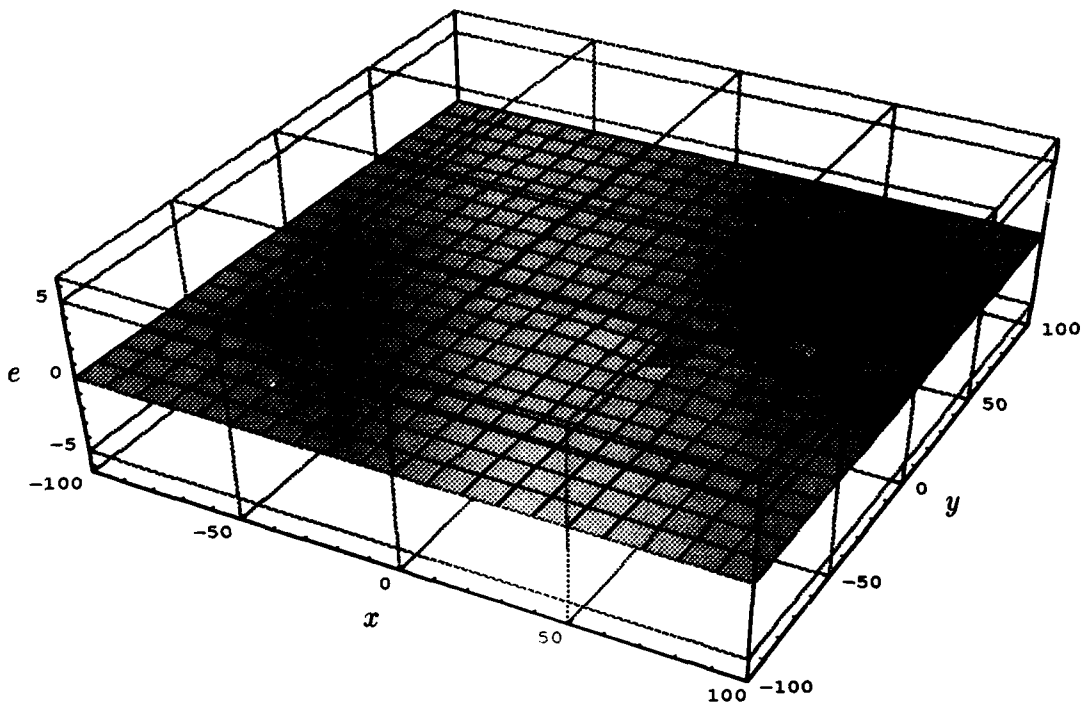


Figure A.12 Bending mode for $\lambda_{12} = -0.24604$.

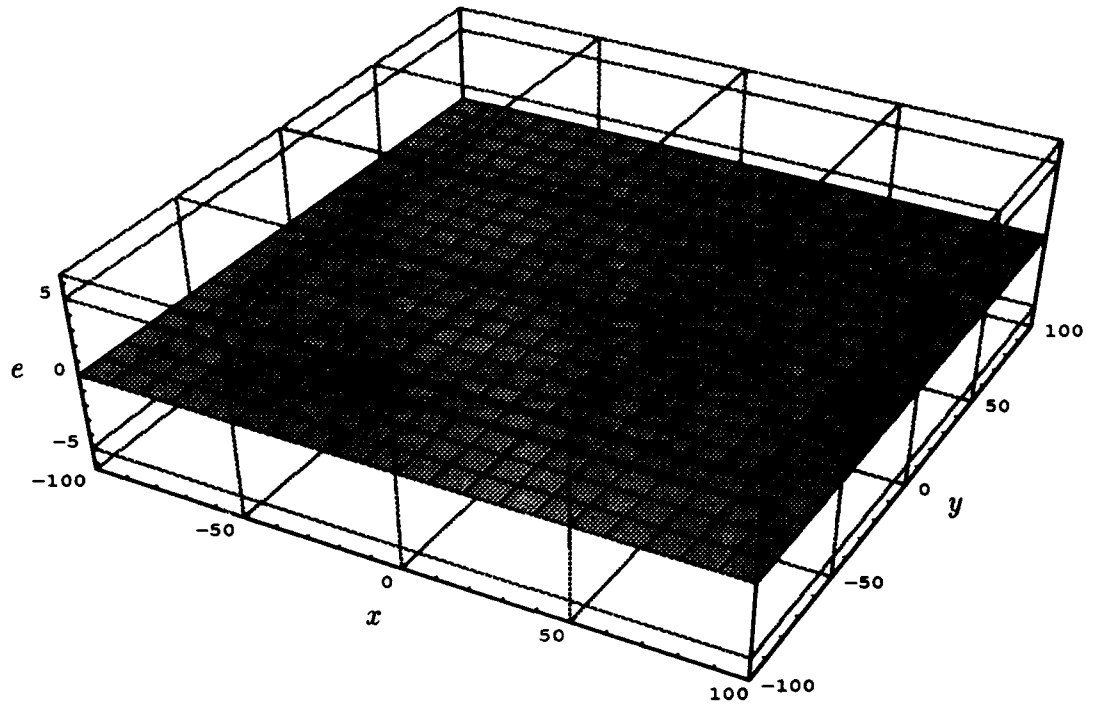


Figure A.13 Bending mode for $\lambda_{13} = -0.24257$.

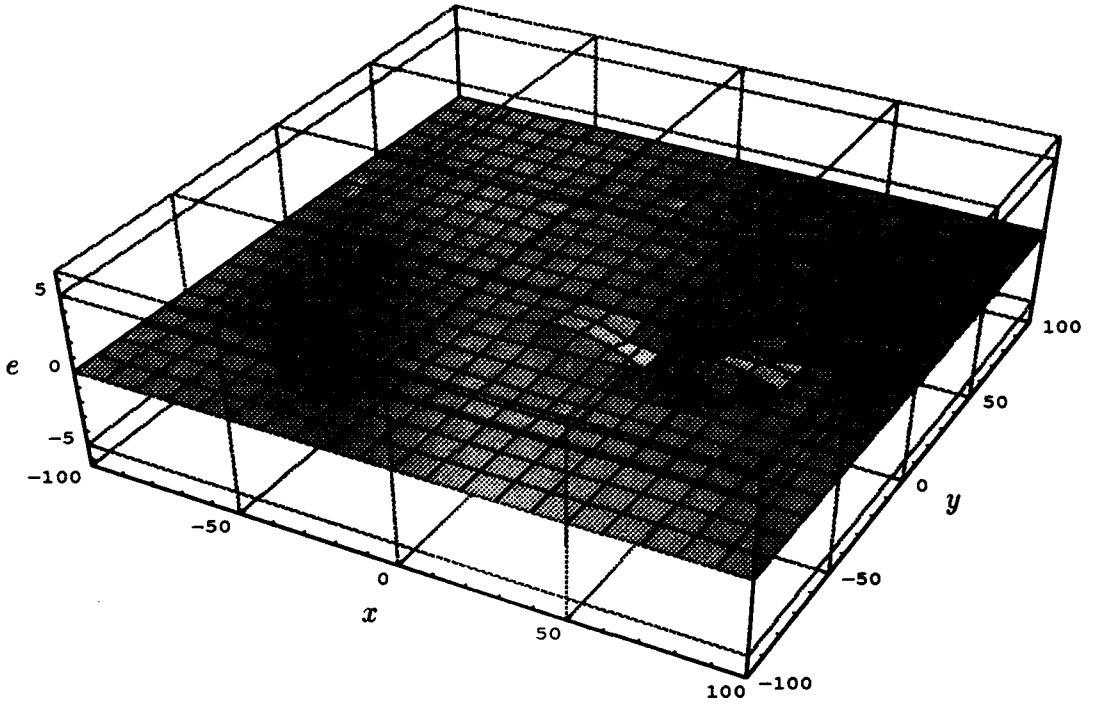


Figure A.14 Bending mode for $\lambda_{14} = -0.22847$.

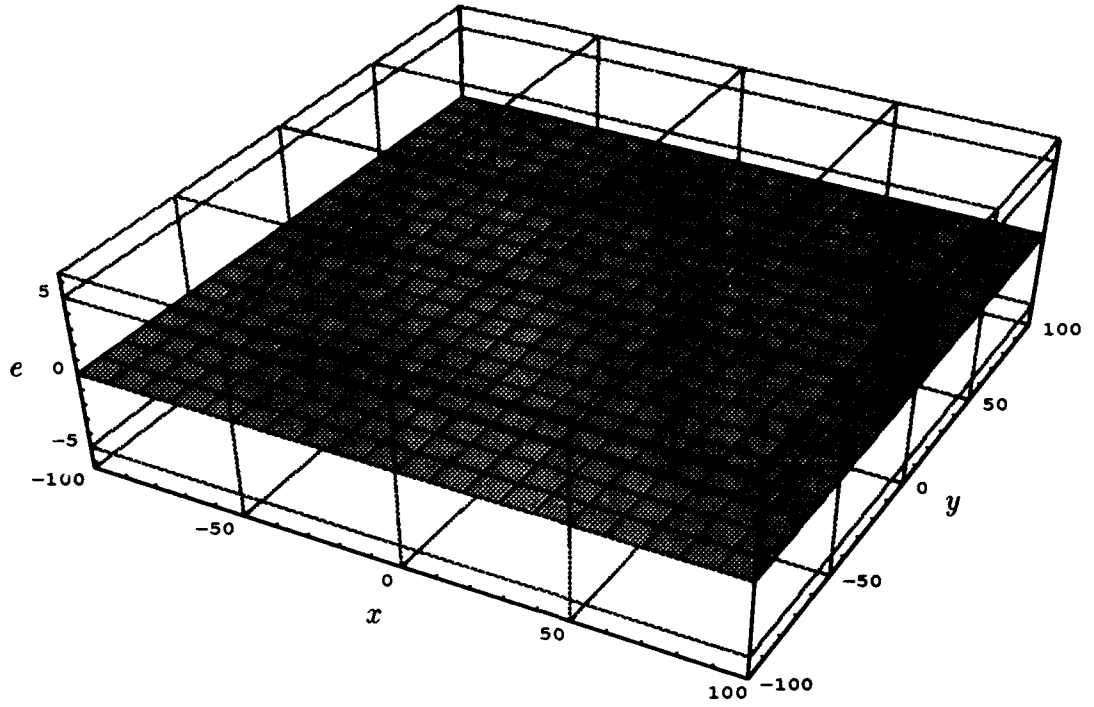


Figure A.15 Bending mode for $\lambda_{15} = -0.22052$.

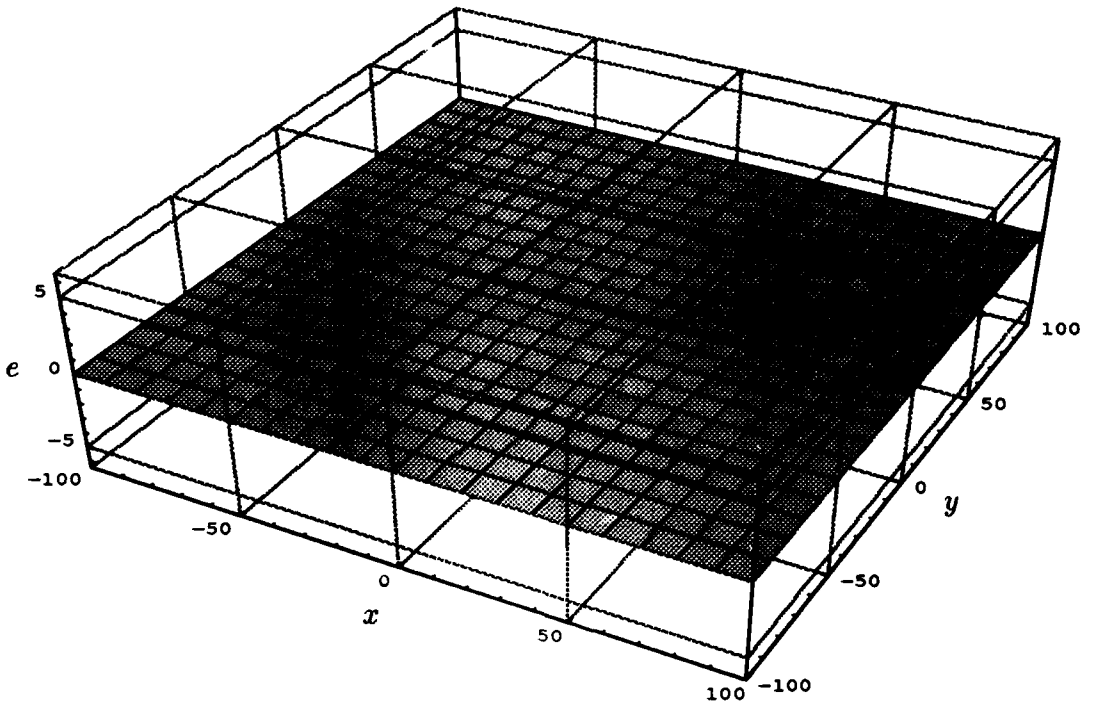


Figure A.16 Bending mode for $\lambda_{16} = -0.21202$.

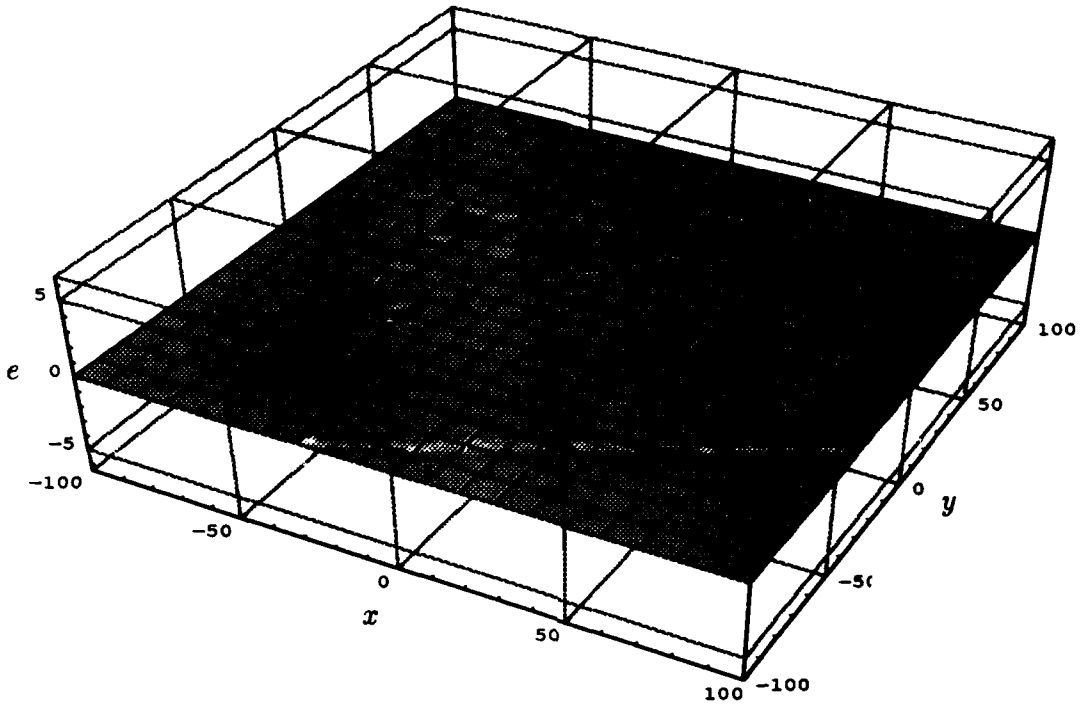


Figure A.17 Bending mode for $\lambda_{17} = -0.20526$.

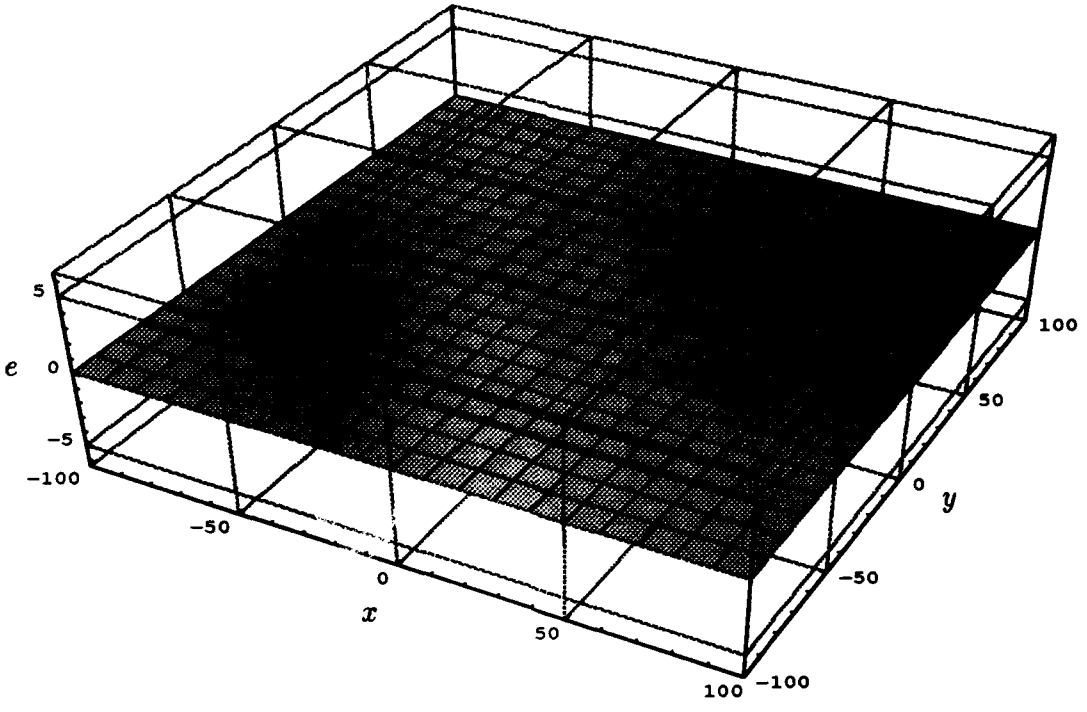


Figure A.18 Bending mode for $\lambda_{18} = -0.20199$.

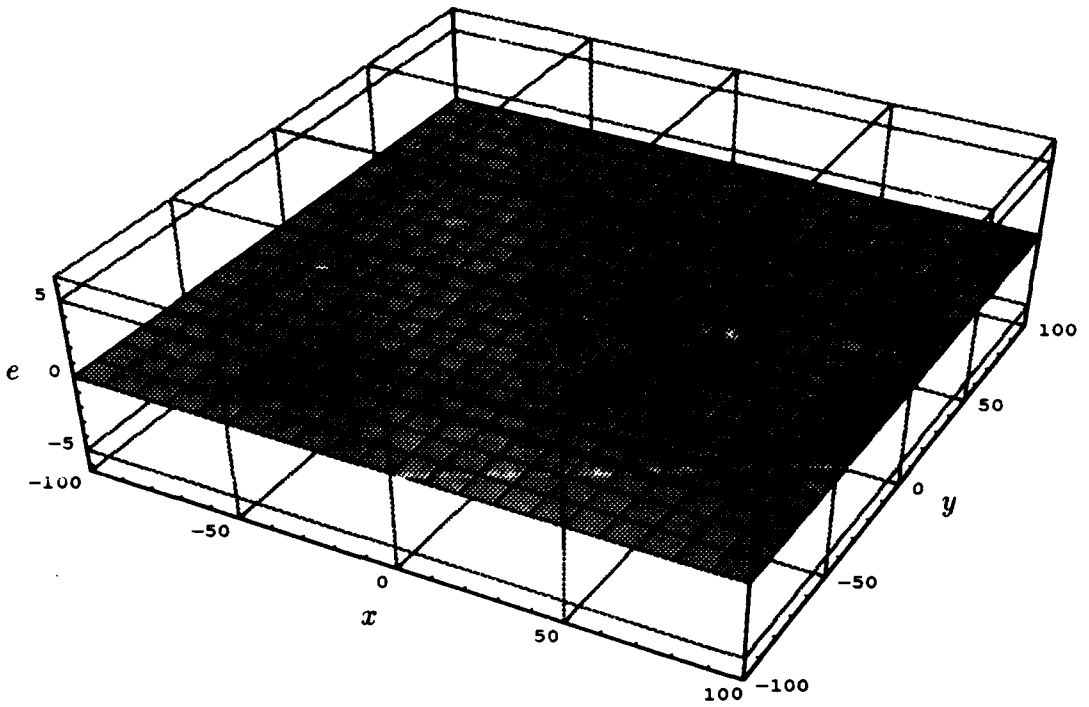


Figure A.19 Bending mode for $\lambda_{19} = -0.16536$.

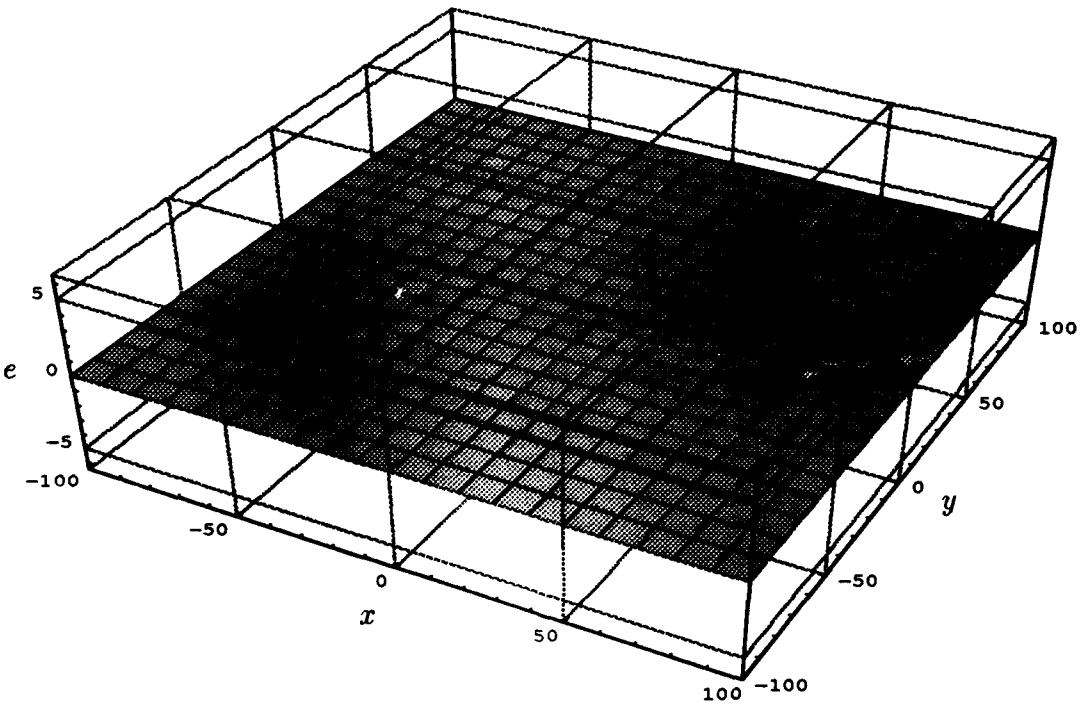


Figure A.20 Bending mode for $\lambda_{20} = -0.15400$.

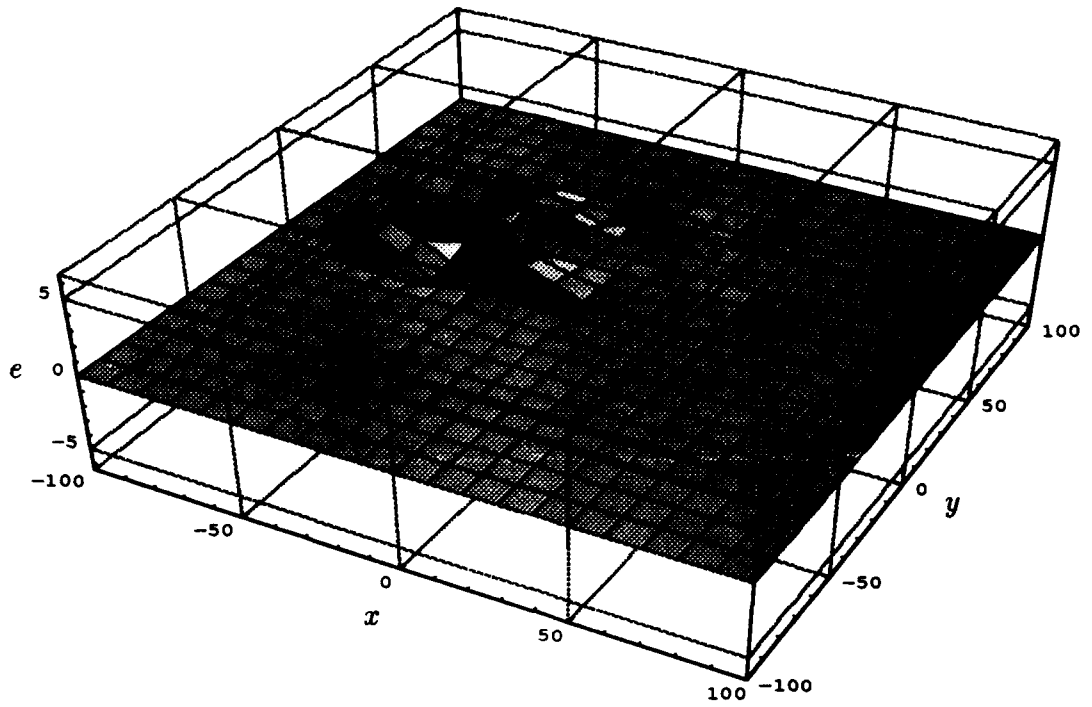


Figure A.21 Bending mode for $\lambda_{21} = -0.14840$.

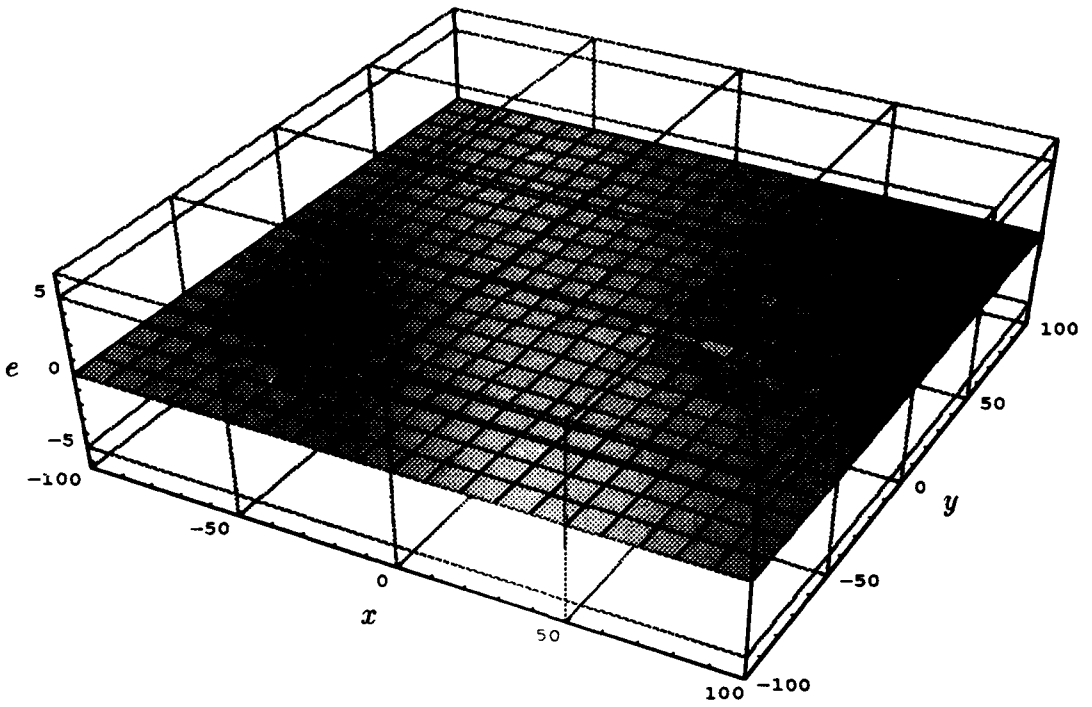


Figure A.22 Bending mode for $\lambda_{22} = -0.12810$.

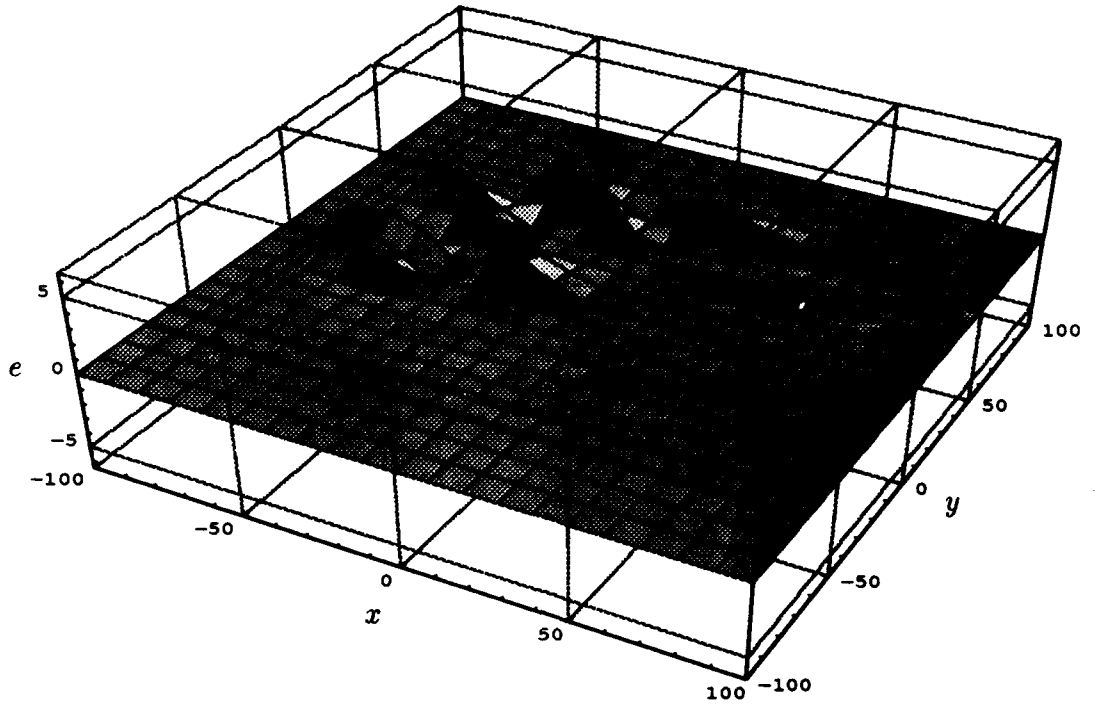


Figure A.23 Bending mode for $\lambda_{23} = -0.12148$.

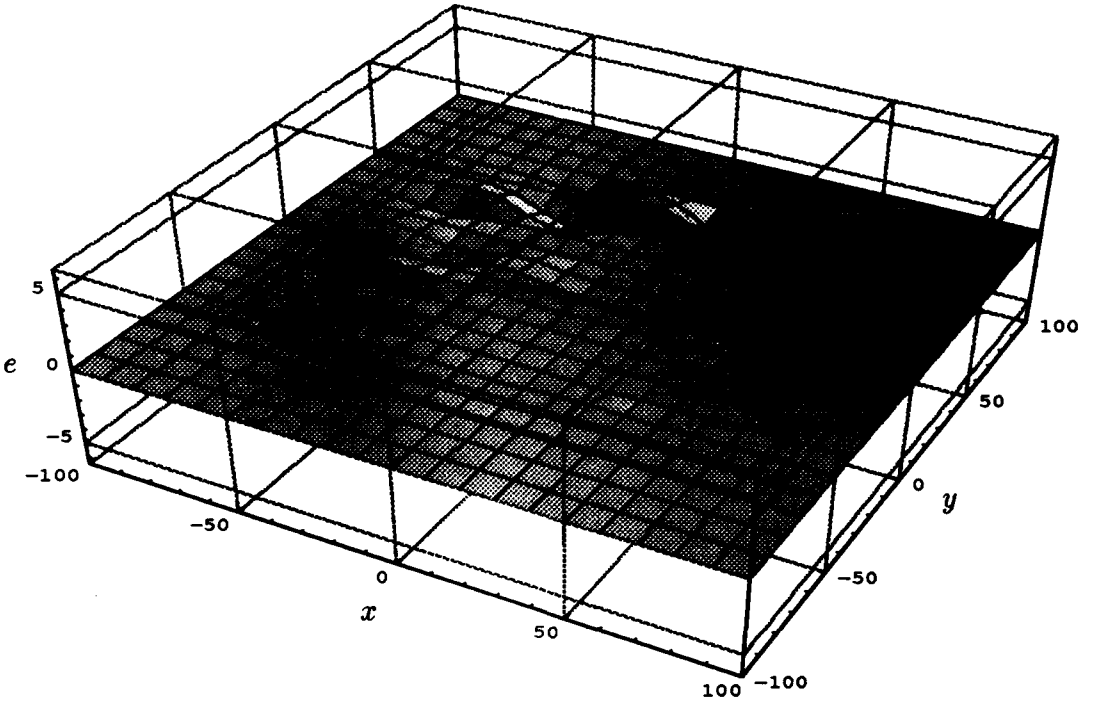


Figure A.24 Bending mode for $\lambda_{24} = -0.11171$.

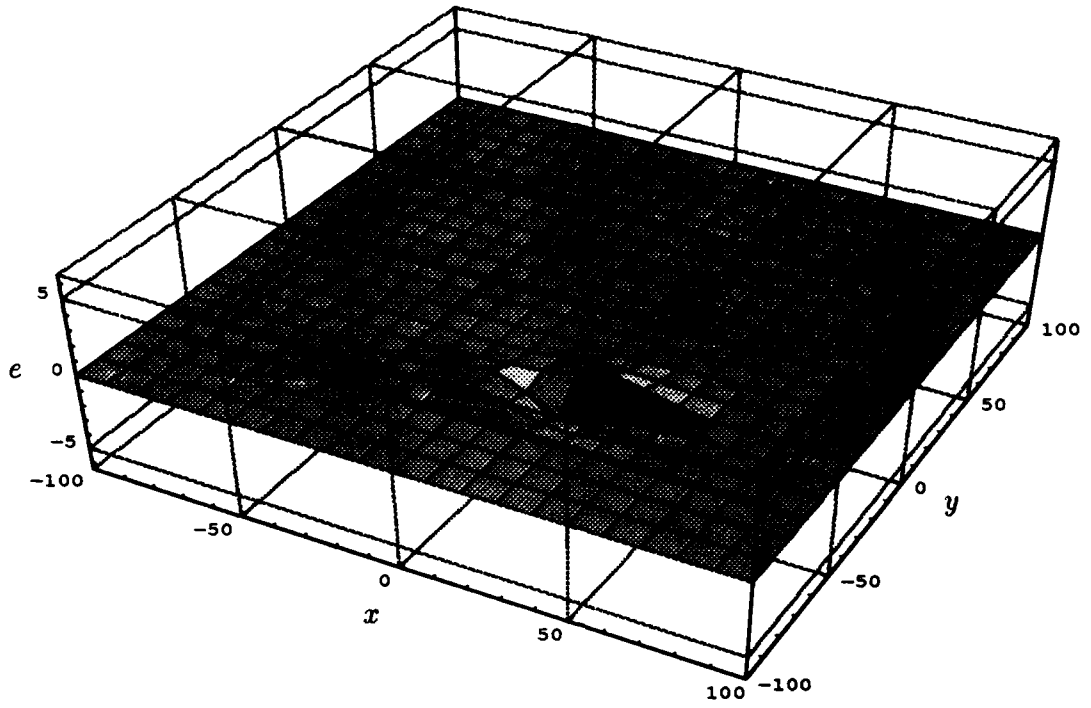


Figure A.25 Bending mode for $\lambda_{25} = -0.10508$.

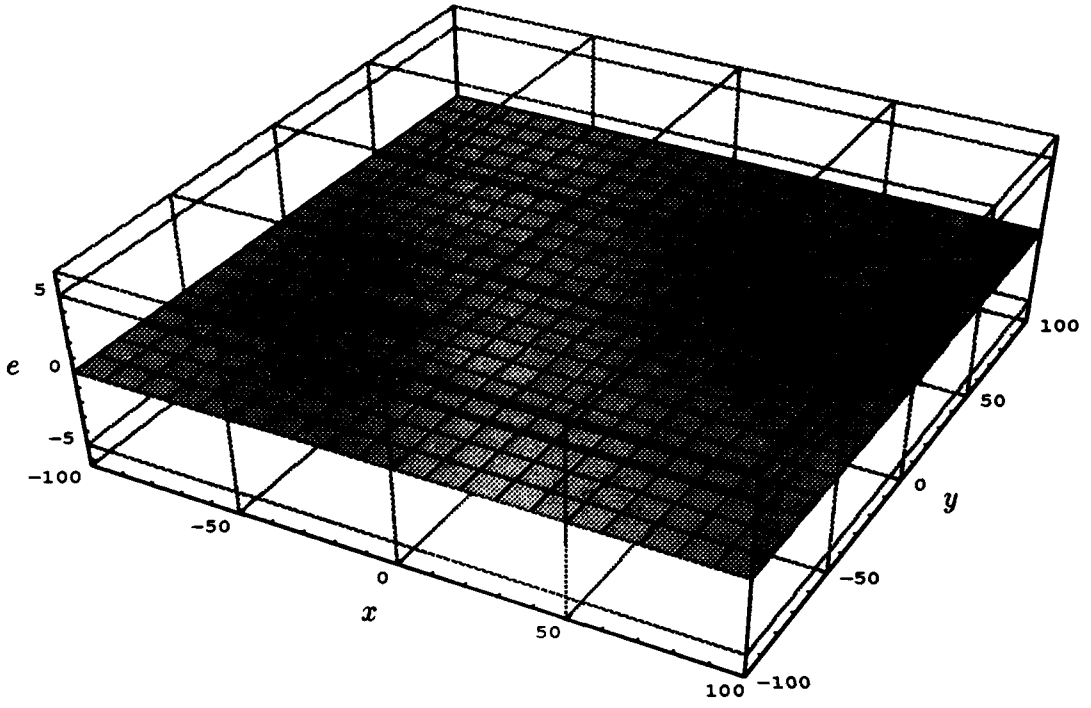


Figure A.26 Bending mode for $\lambda_{26} = -0.094732$.

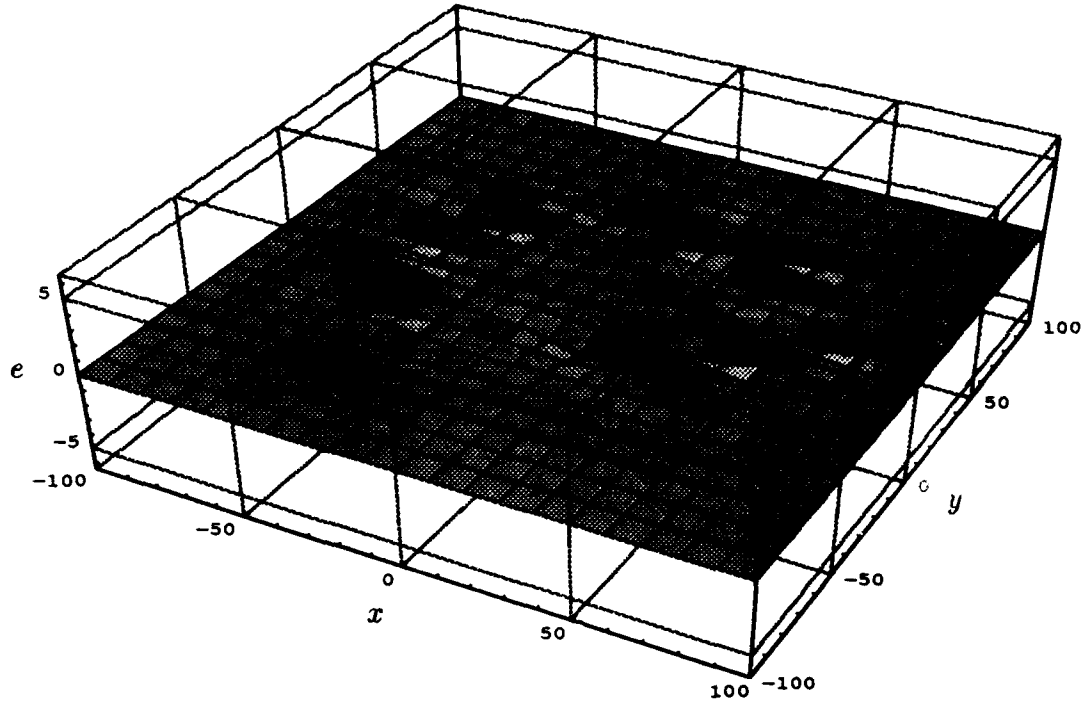


Figure A.27 Bending mode for $\lambda_{27} = -0.088829$.

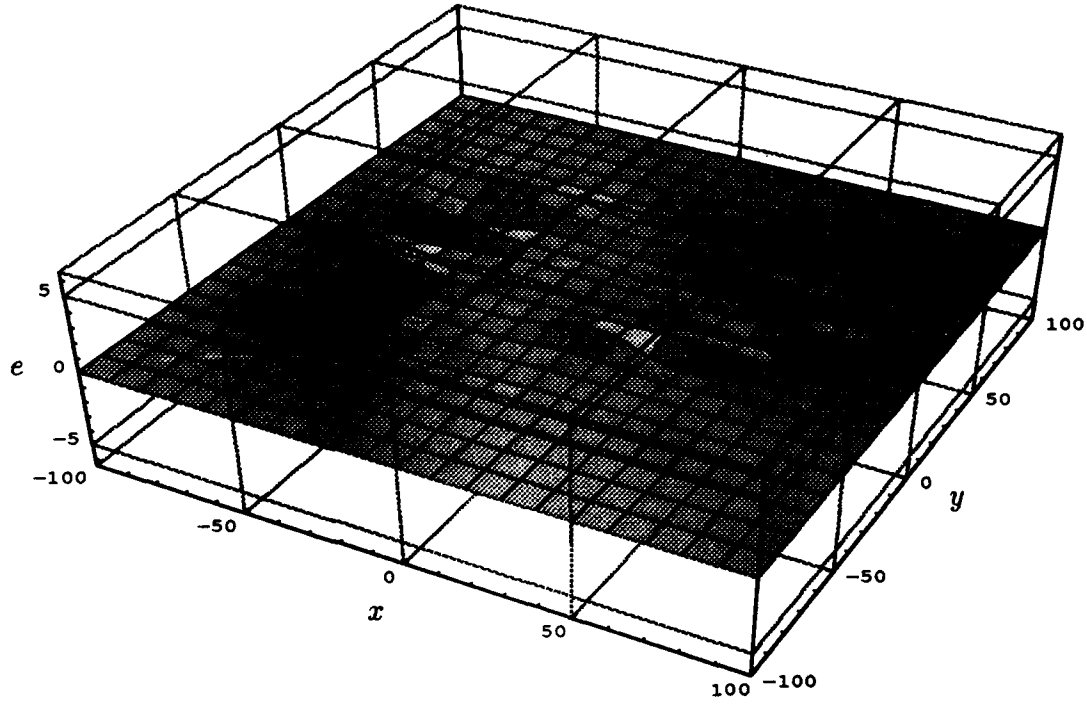


Figure A.28 Bending mode for $\lambda_{28} = -0.087321$.

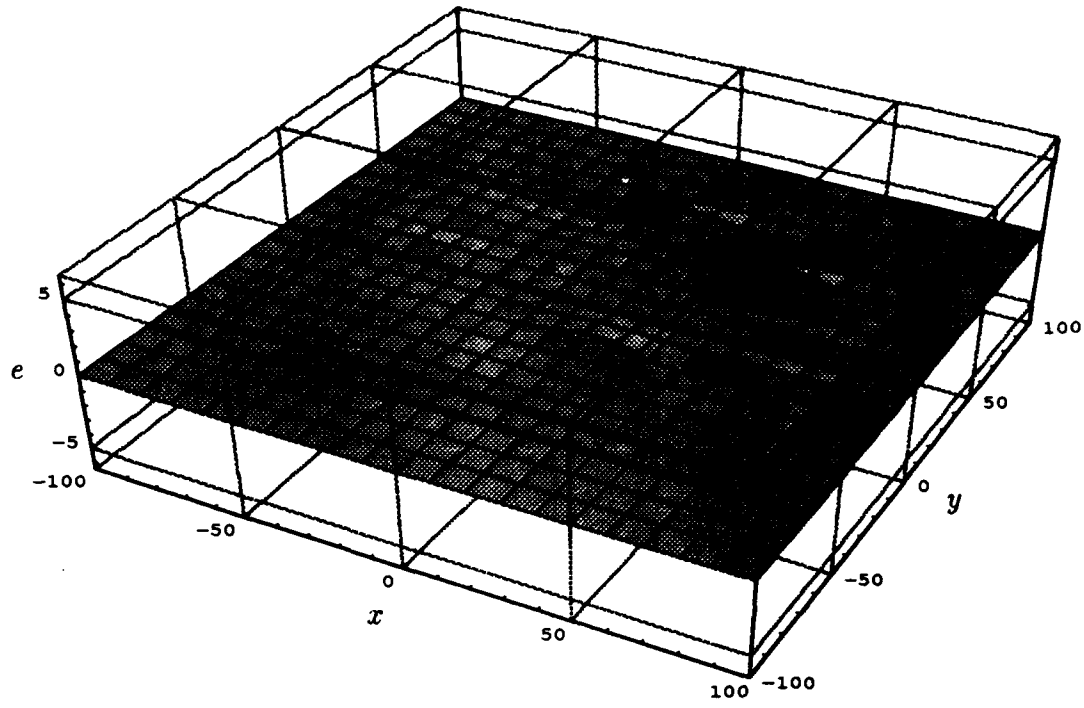


Figure A.29 Bending mode for $\lambda_{29} = -0.080692$.

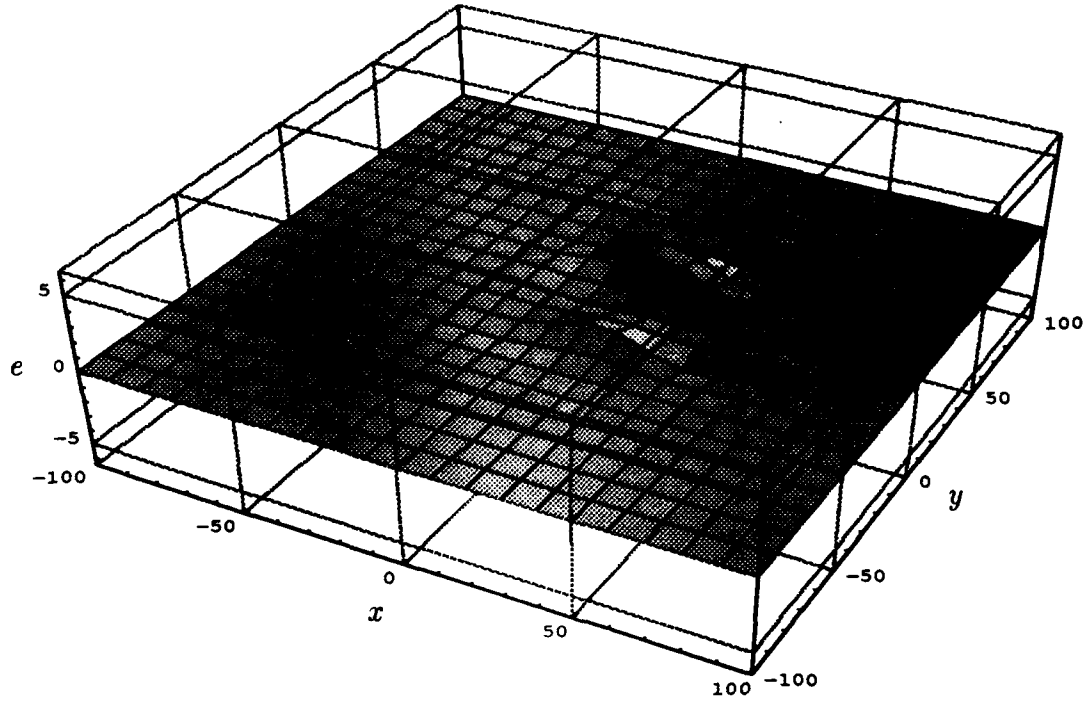


Figure A.30 Bending mode for $\lambda_{30} = -0.076363$.

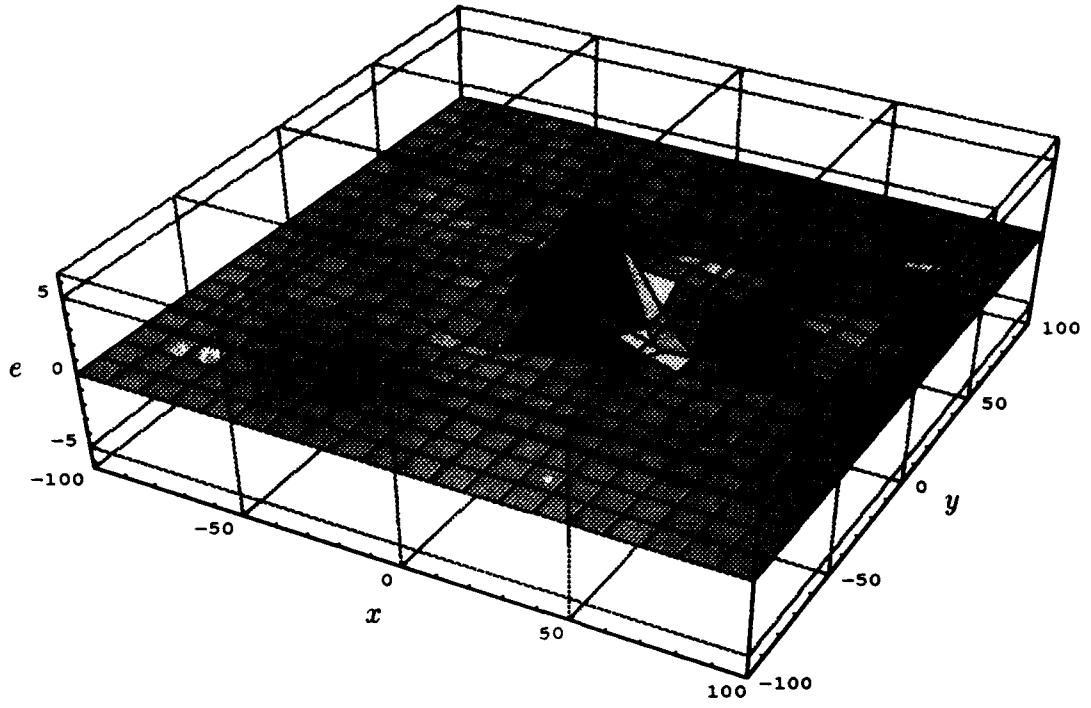


Figure A.31 Bending mode for $\lambda_{31} = -0.075163$.

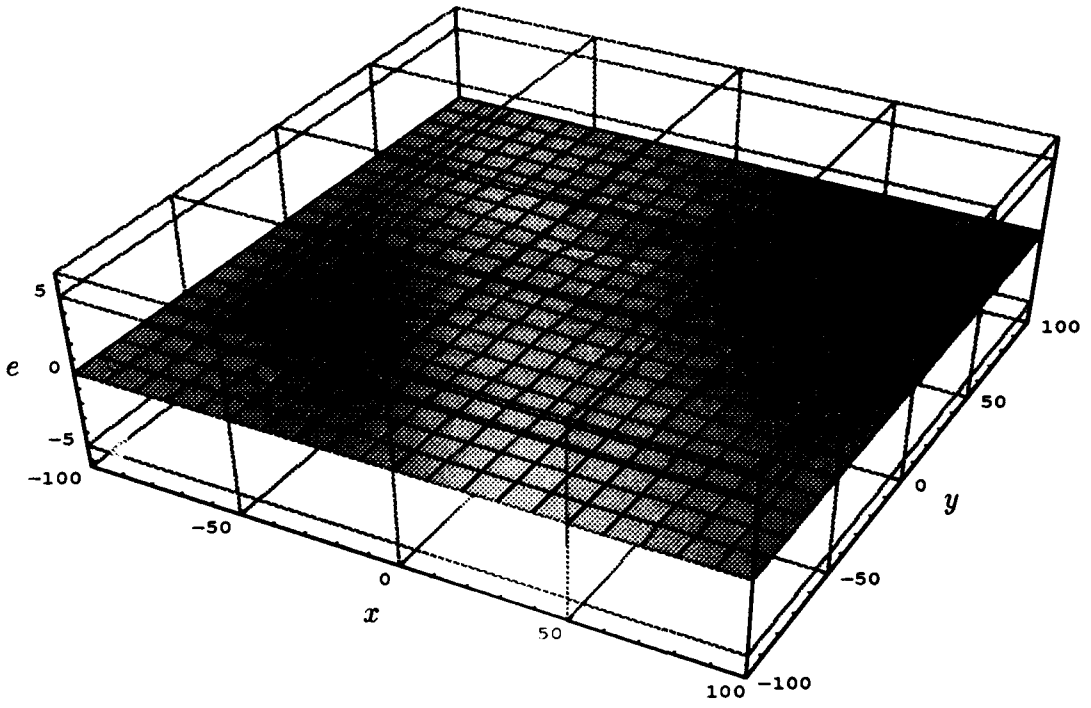


Figure A.32 Bending mode for $\lambda_{32} = -0.068943$.

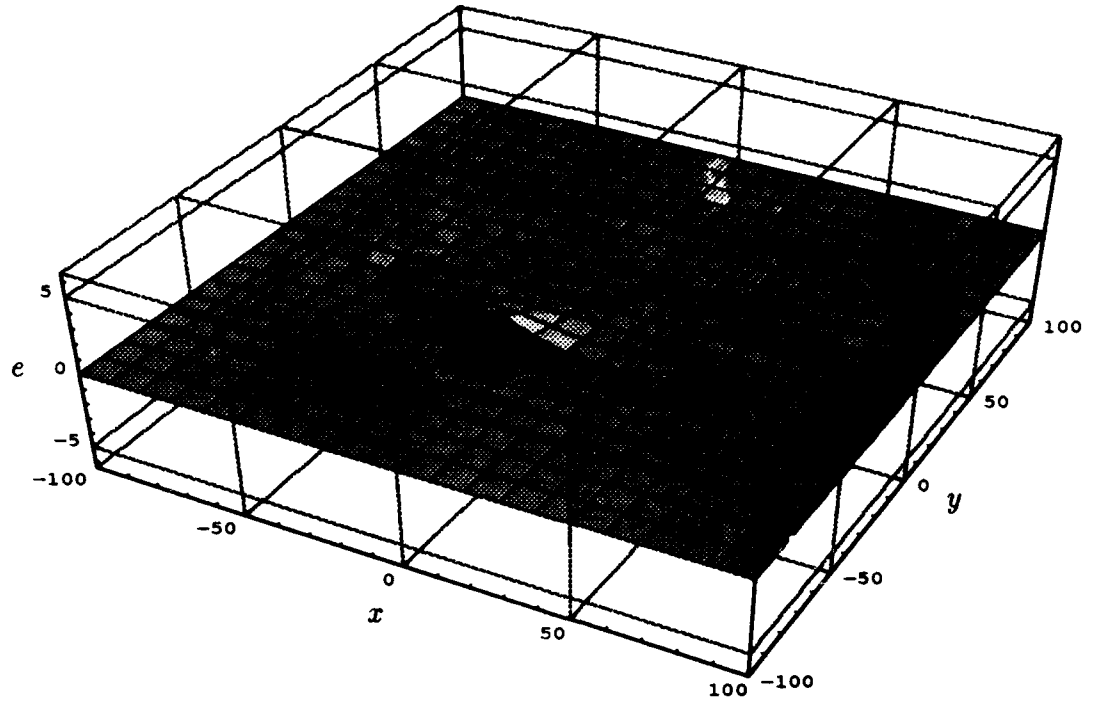


Figure A.33 Bending mode for $\lambda_{33} = -0.068699$.

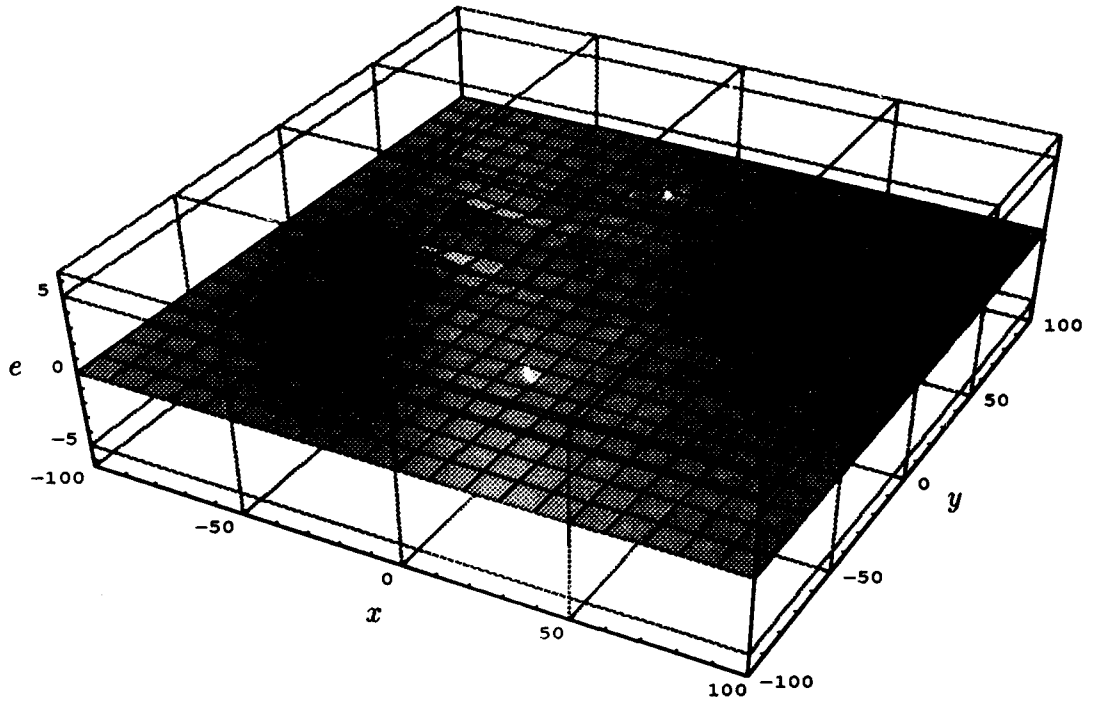


Figure A.34 Bending mode for $\lambda_{34} = -0.068205$.

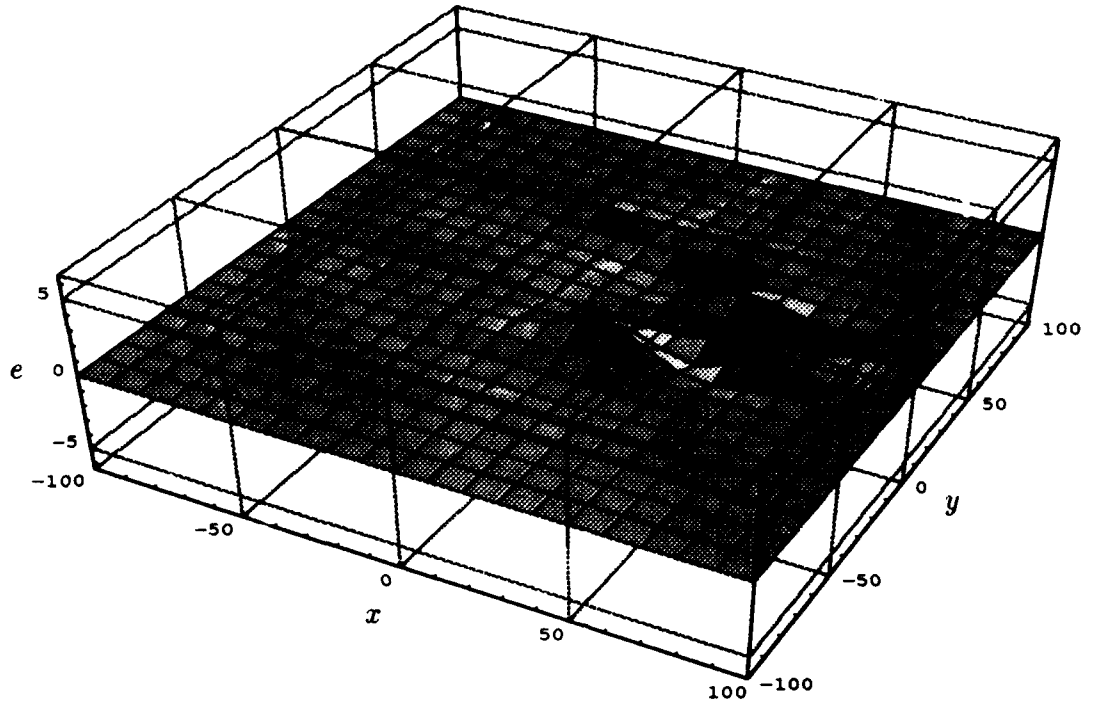


Figure A.35 Bending mode for $\lambda_{35} = -0.065954$.

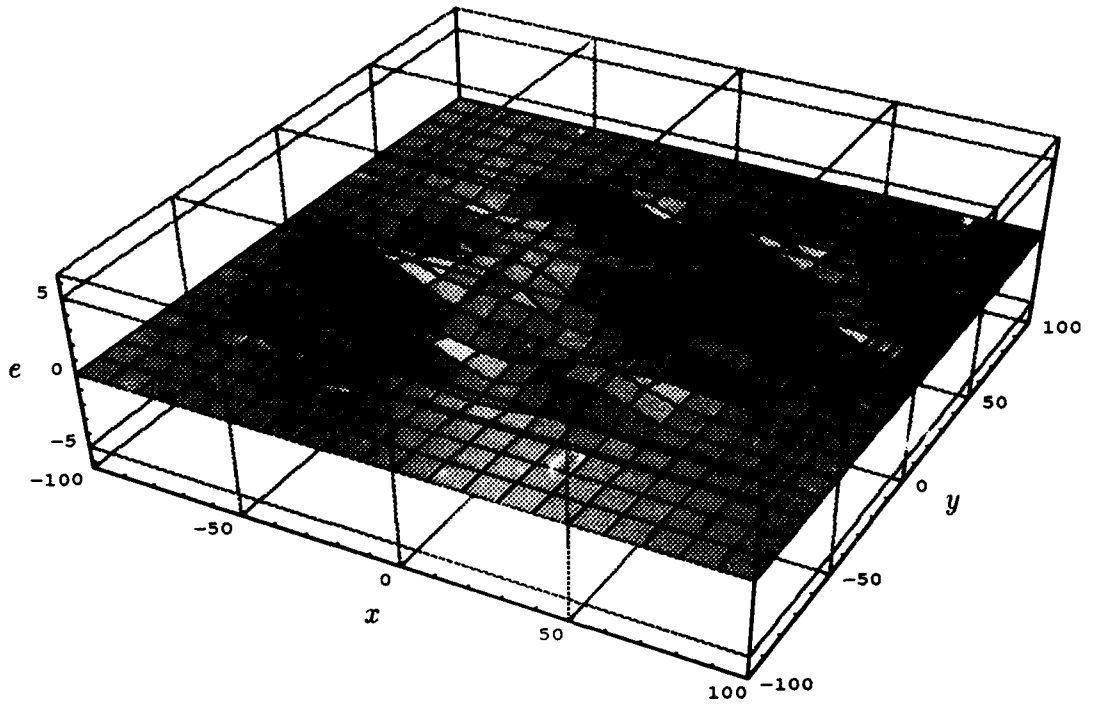


Figure A.36 Bending mode for $\lambda_{36} = -0.061600$.

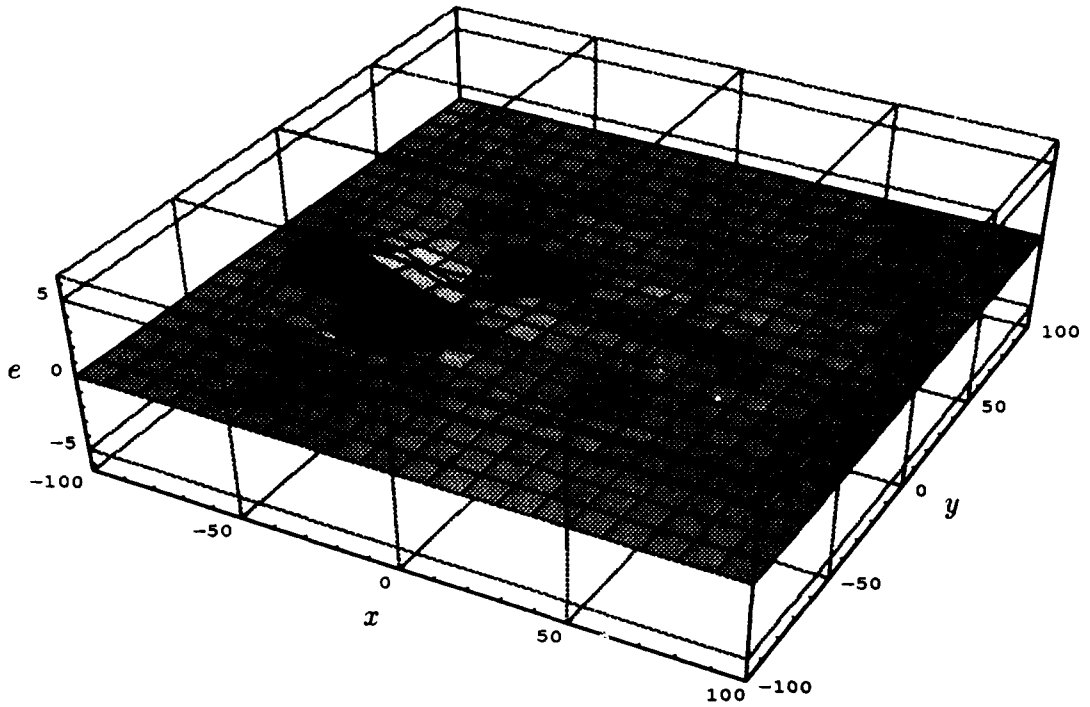


Figure A.37 Bending mode for $\lambda_{37} = -0.058455$.

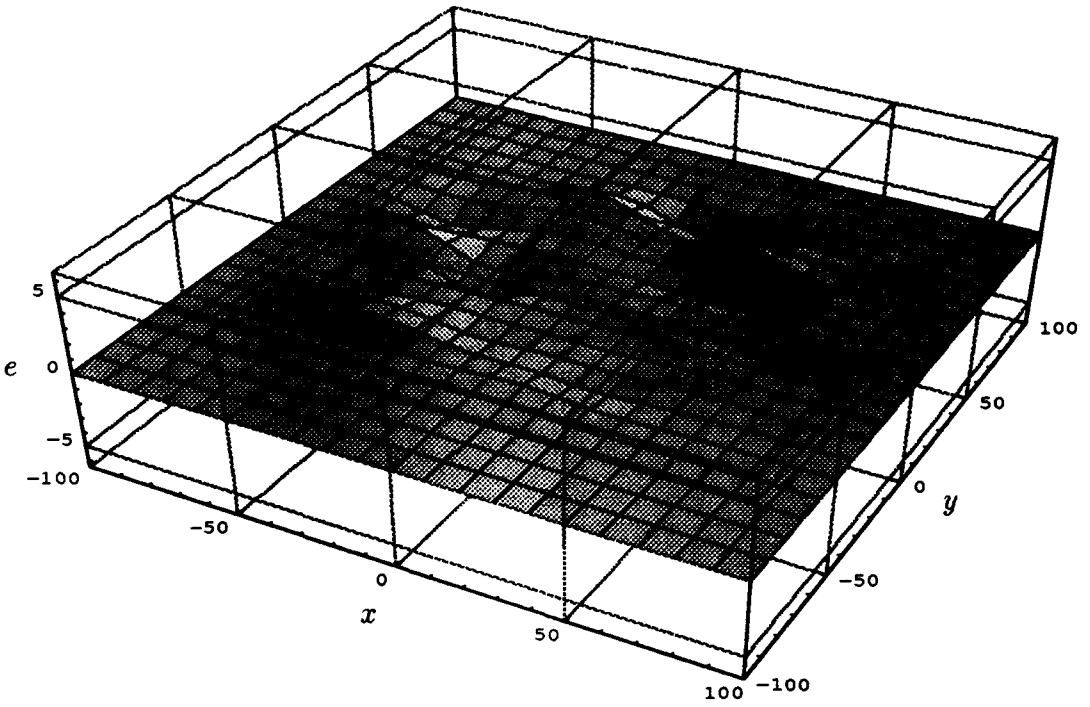


Figure A.38 Bending mode for $\lambda_{38} = -0.055670$.

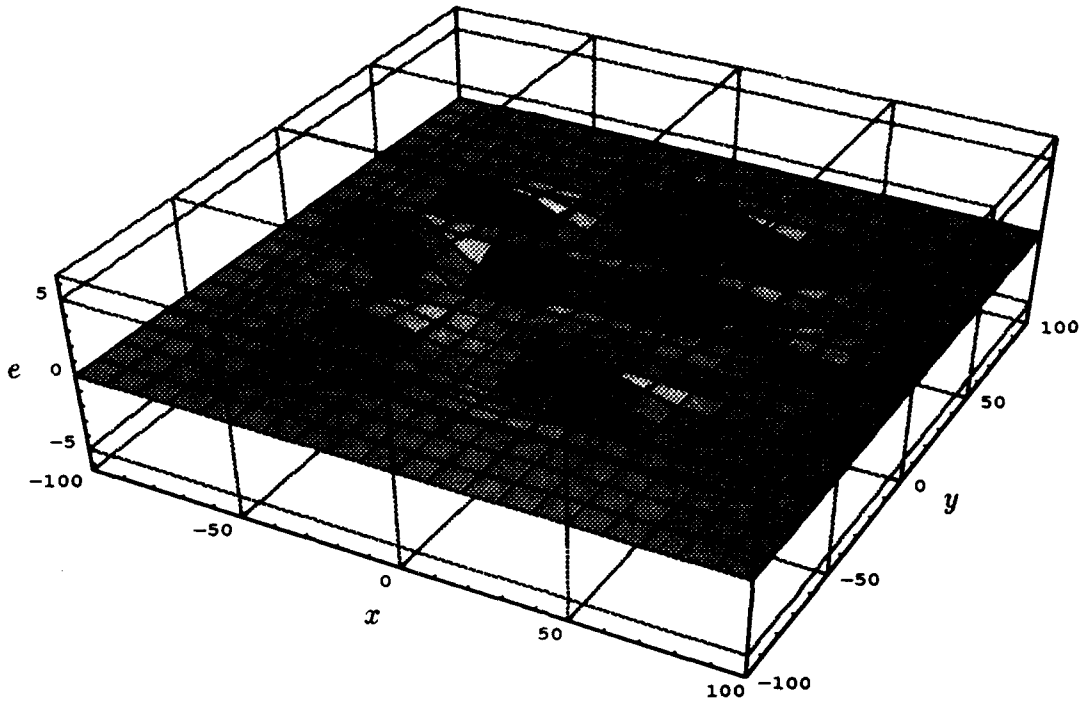


Figure A.39 Bending mode for $\lambda_{39} = -0.054397$.

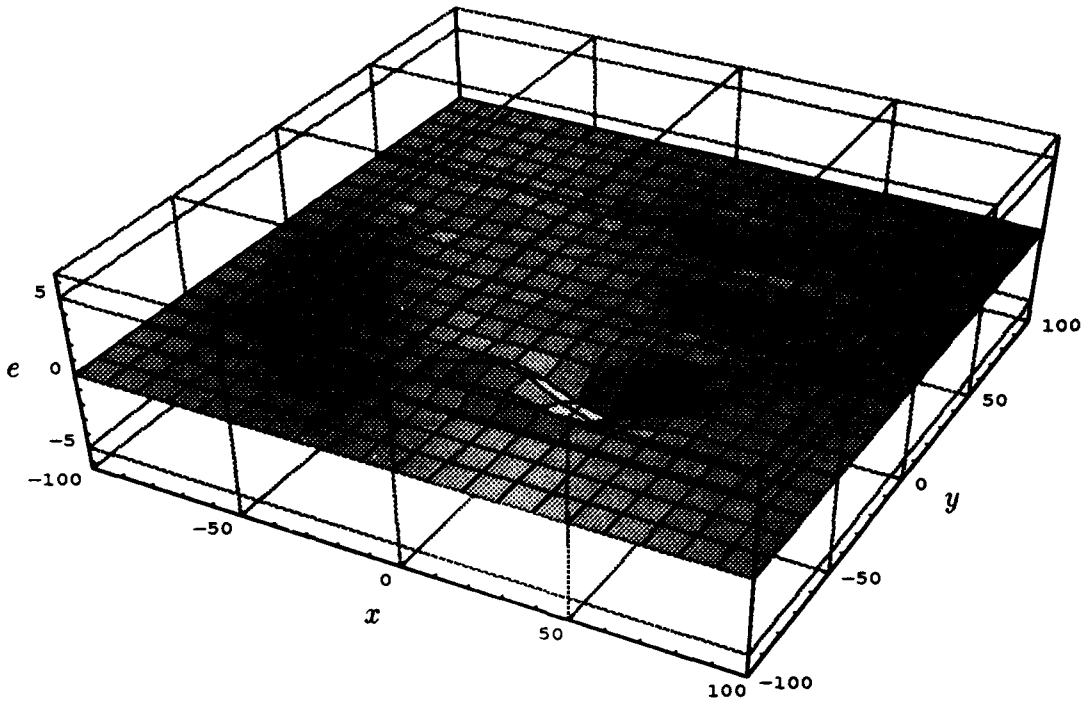


Figure A.40 Bending mode for $\lambda_{40} = -0.052854$.

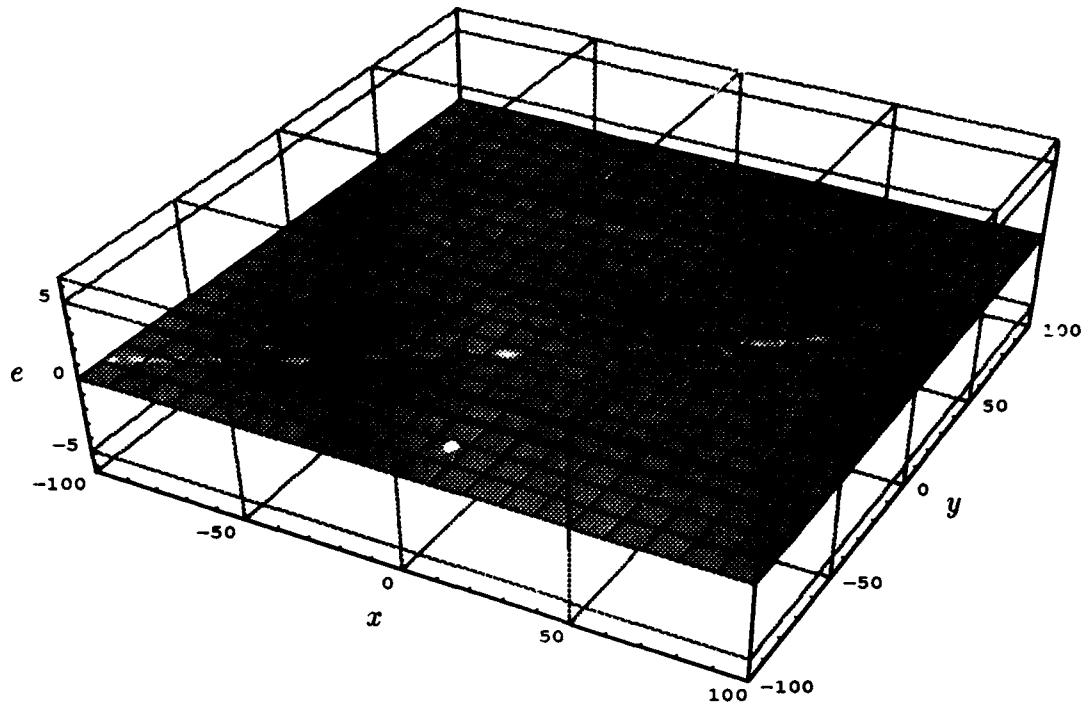


Figure A.41 Bending mode for $\lambda_{41} = -0.048476$.

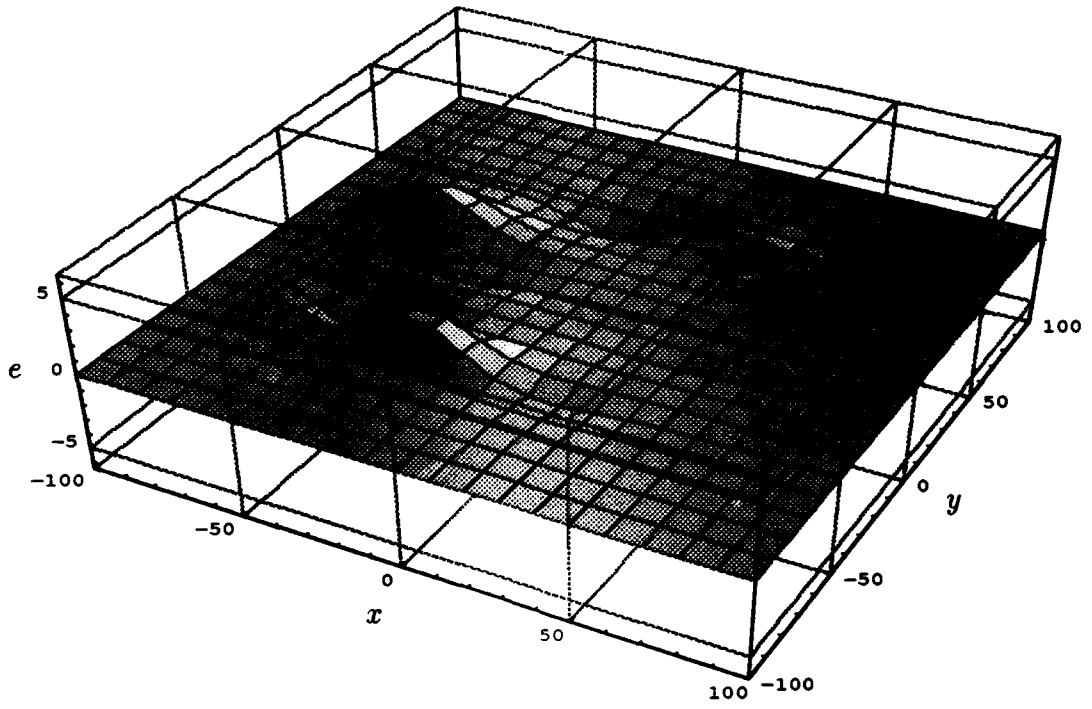


Figure A.42 Bending mode for $\lambda_{42} = -0.046709$.

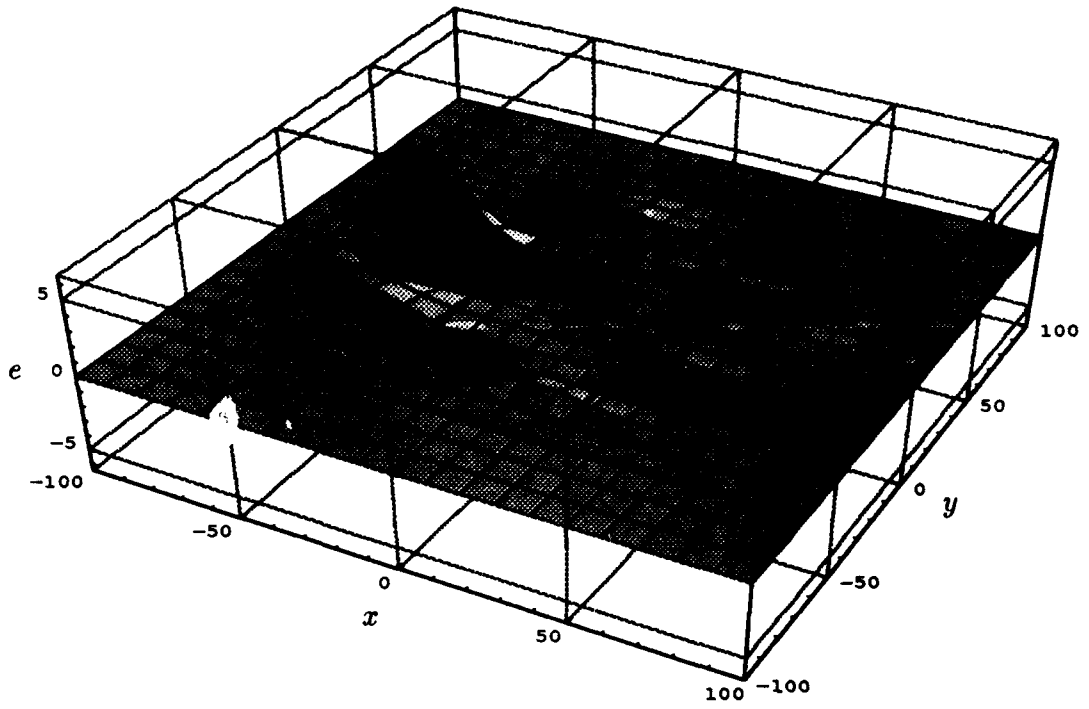


Figure A.43 Bending mode for $\lambda_{43} = -0.045590$.

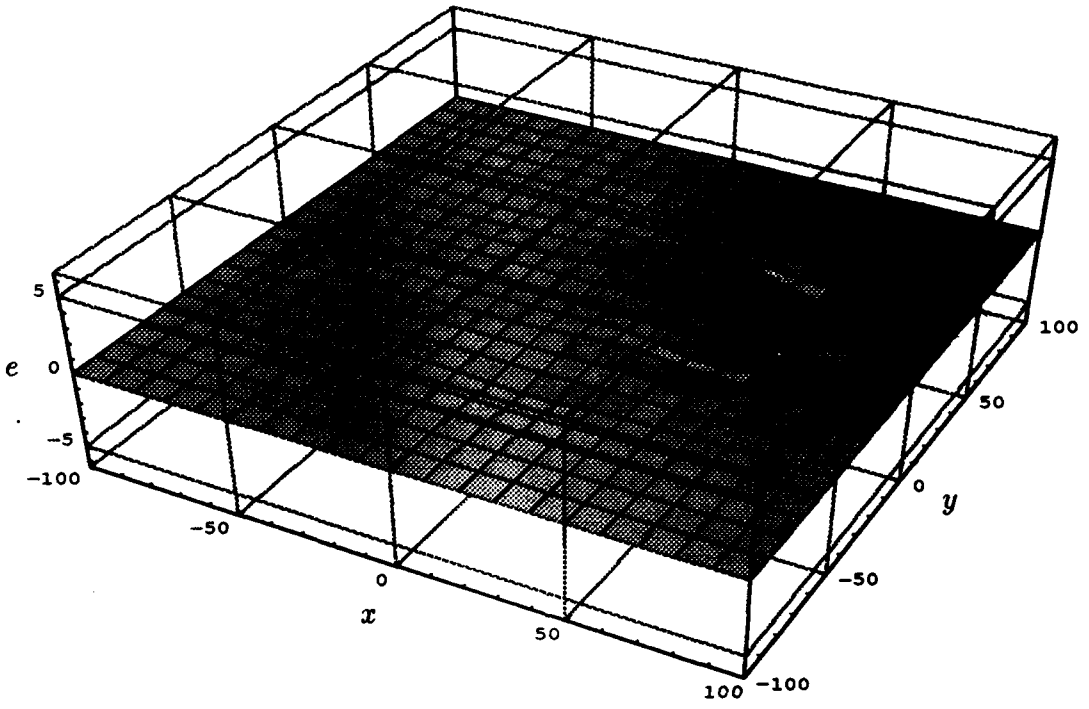


Figure A.44 Bending mode for $\lambda_{44} = -0.039550$.

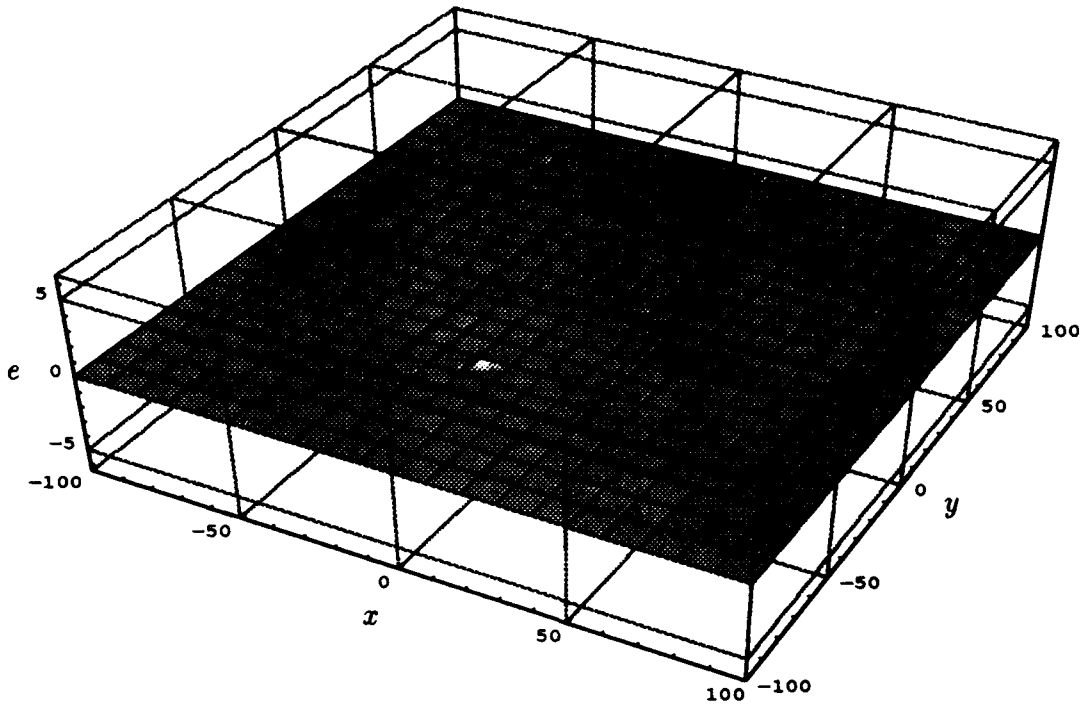


Figure A.45 Bending mode for $\lambda_{45} = -0.035200$.

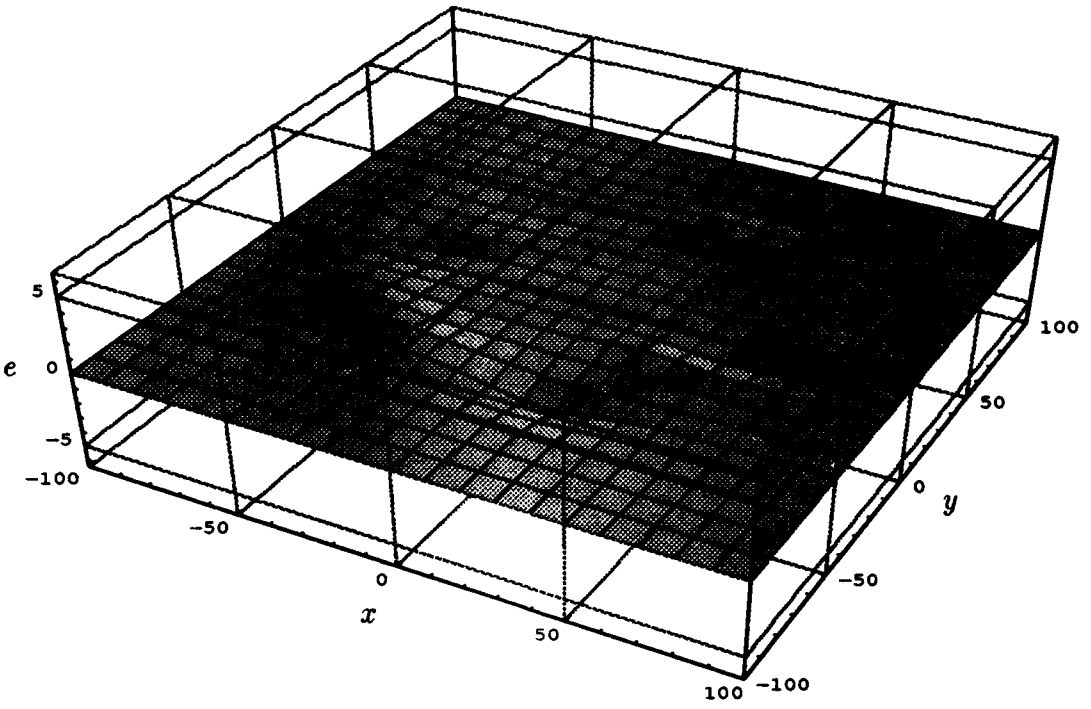


Figure A.46 Bending mode for $\lambda_{46} = -0.033180$.

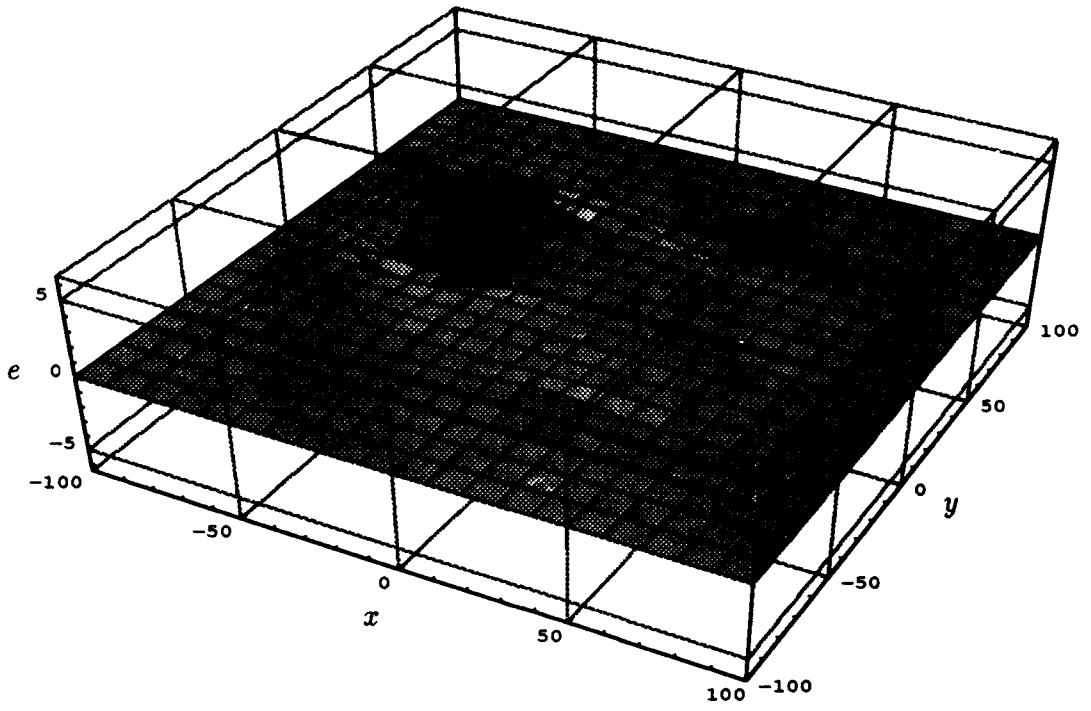


Figure A.47 Bending mode for $\lambda_{47} = -0.032208$.

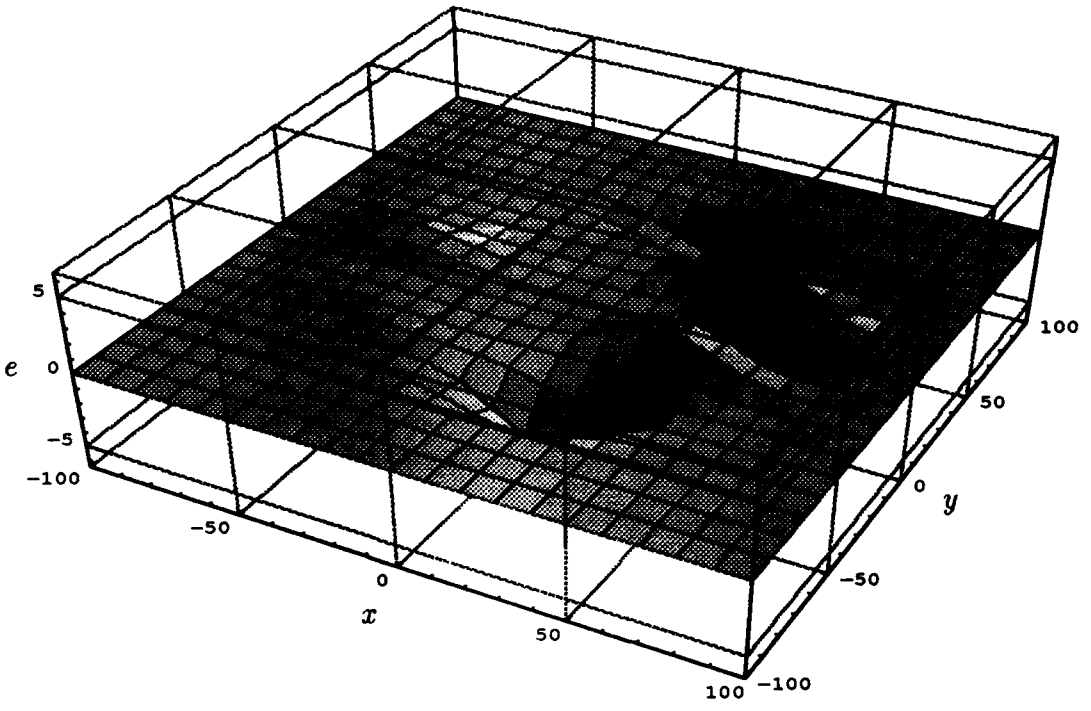


Figure A.48 Bending mode for $\lambda_{48} = -0.028752$.

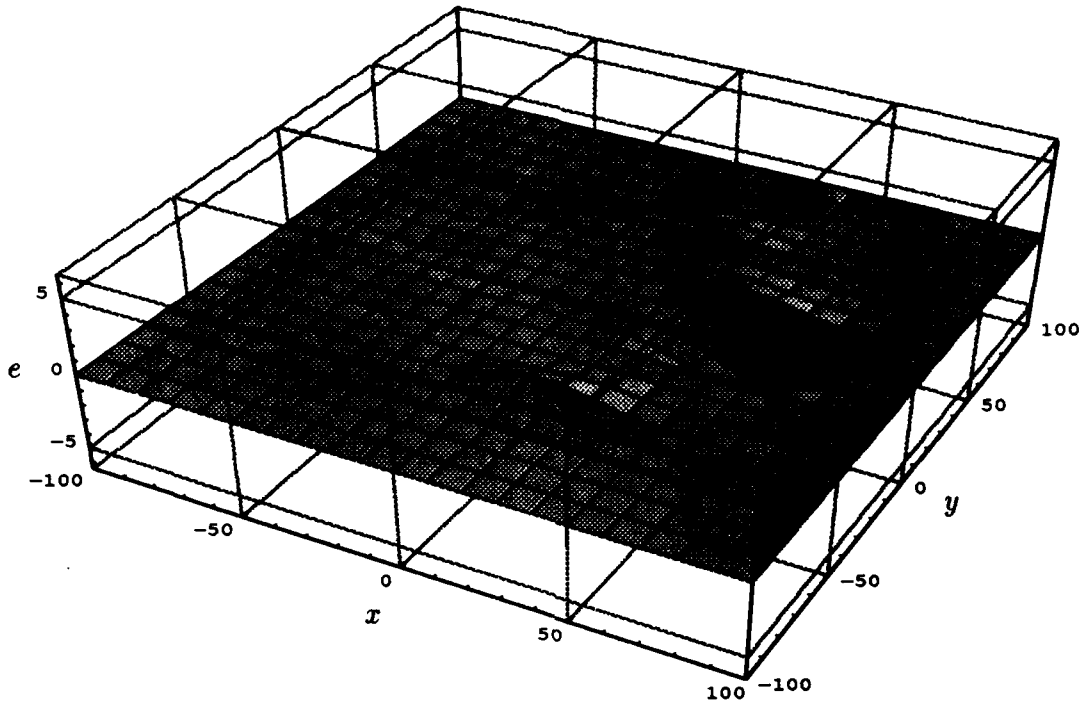


Figure A.49 Bending mode for $\lambda_{49} = -0.022311$.

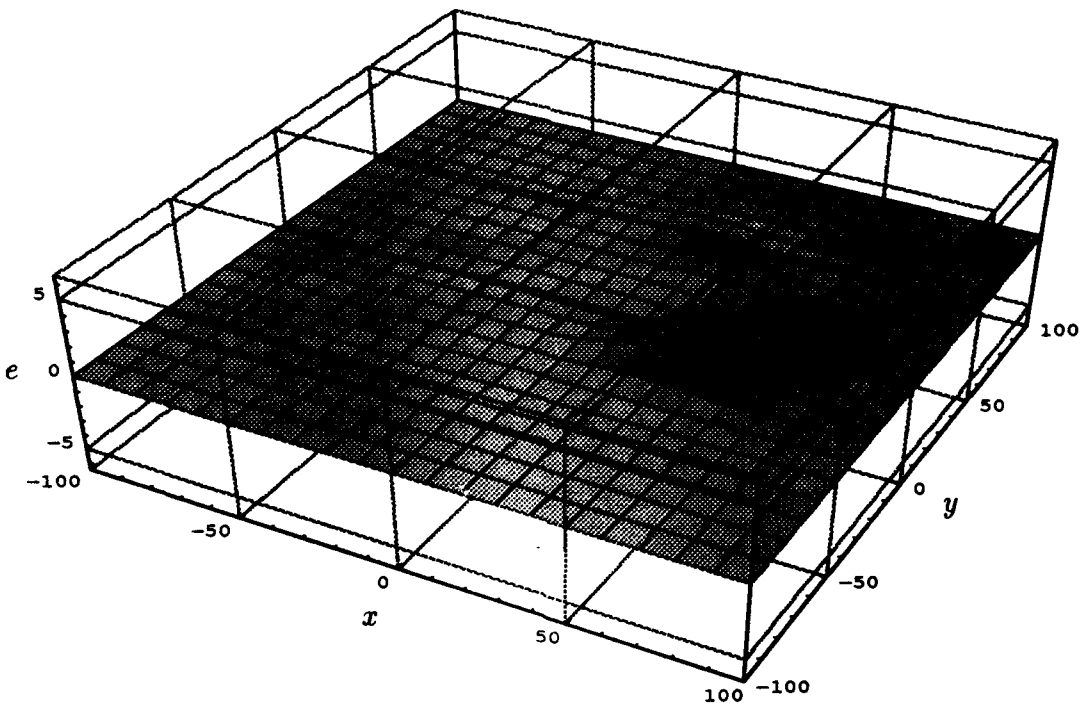


Figure A.50 Bending mode for $\lambda_{50} = -0.021602$.

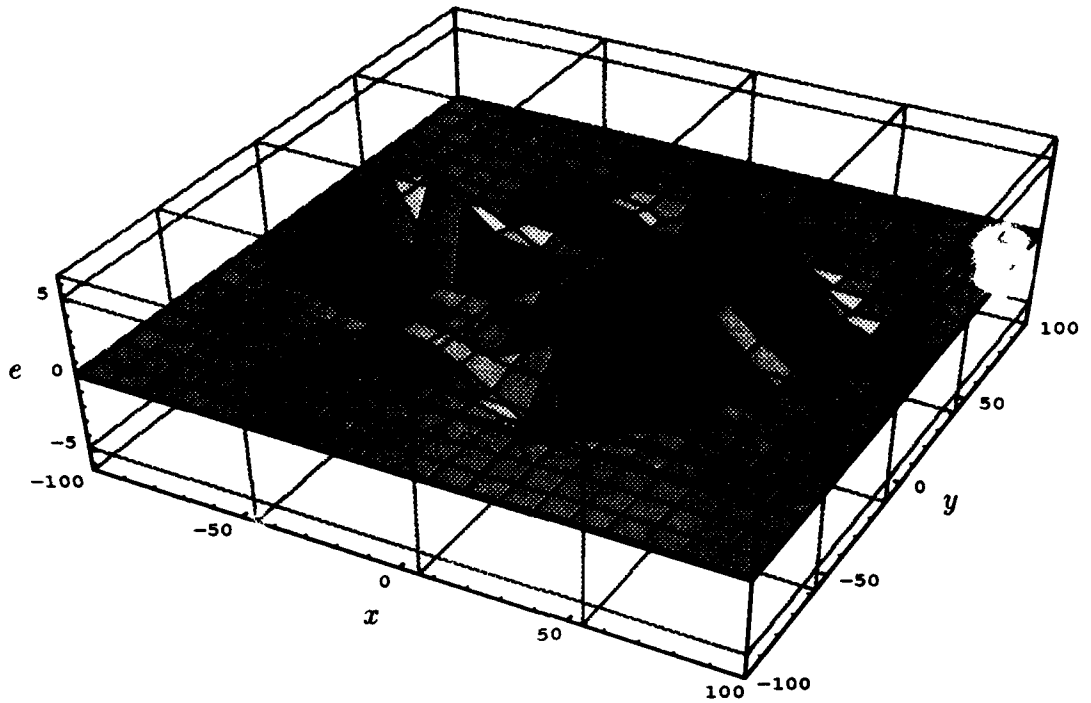


Figure A.51 Sum of all fifty bending modes.

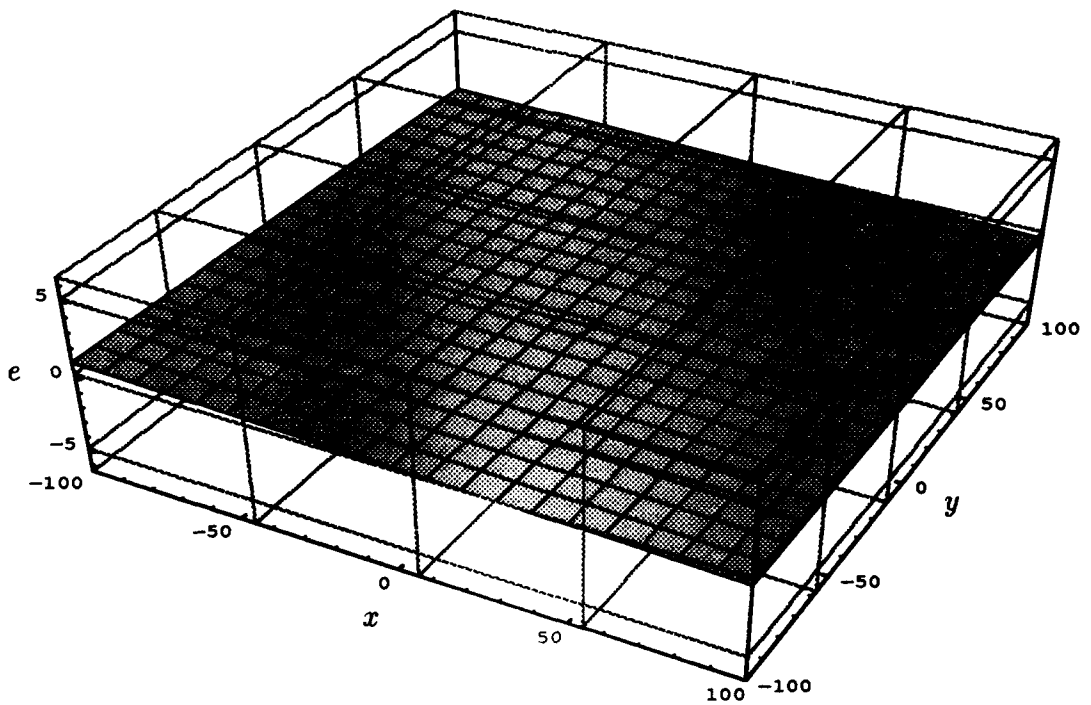


Figure A.52 Regression plane which all bending modes are bent 'from'.

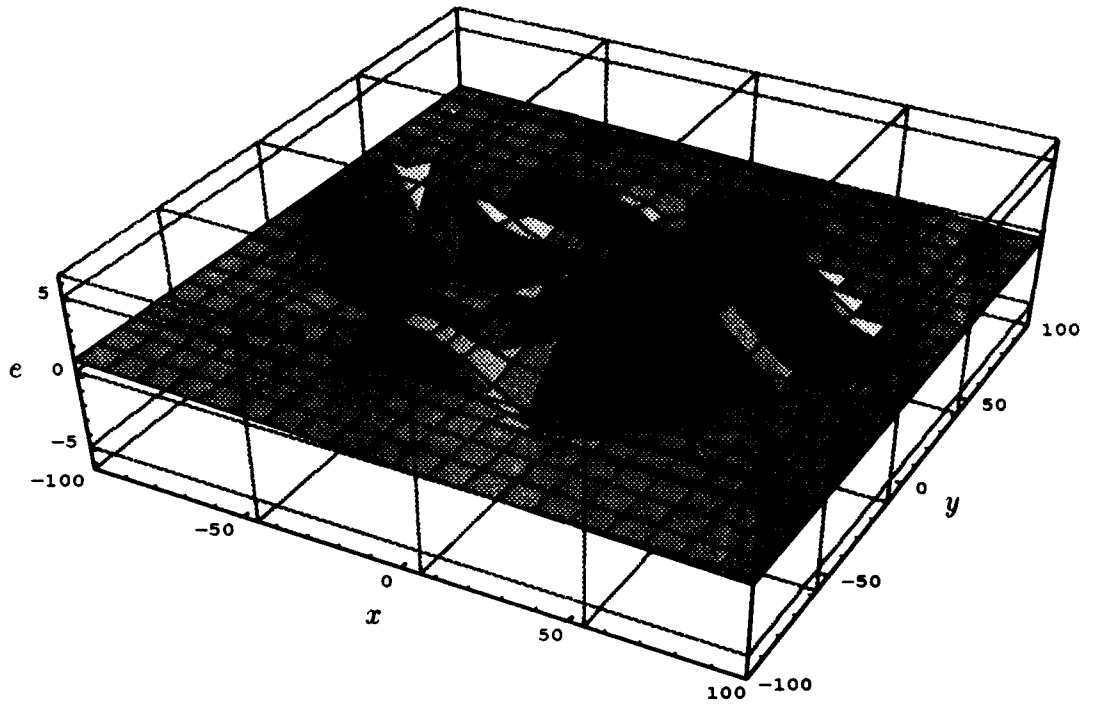


Figure A.53 Sum Bending Modes and Regression: Interpolation and extrapolation of residuals.

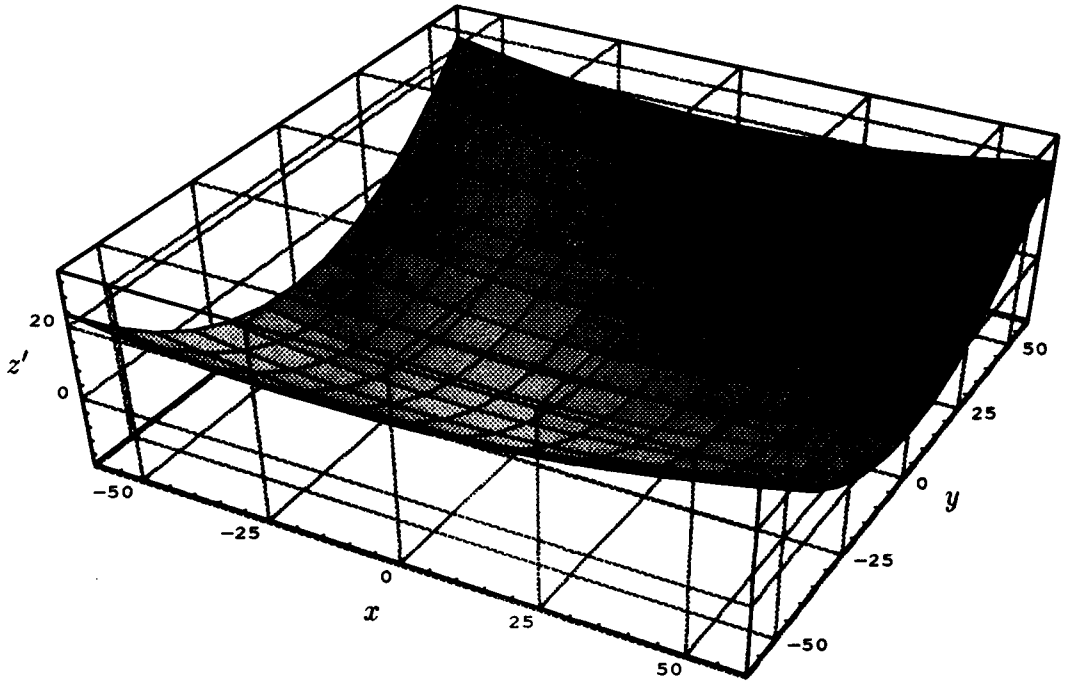


Figure A.54 Polynomial subtracted from physical z coordinates to form residuals.

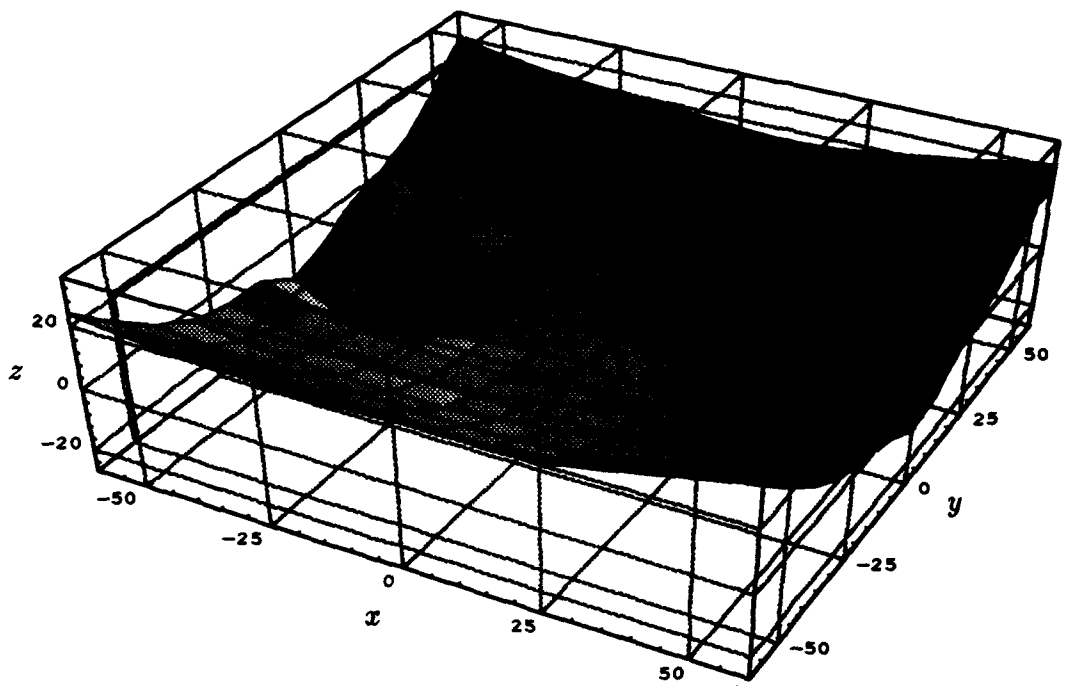


Figure A.55 Estimated height of face in physical dimensions.

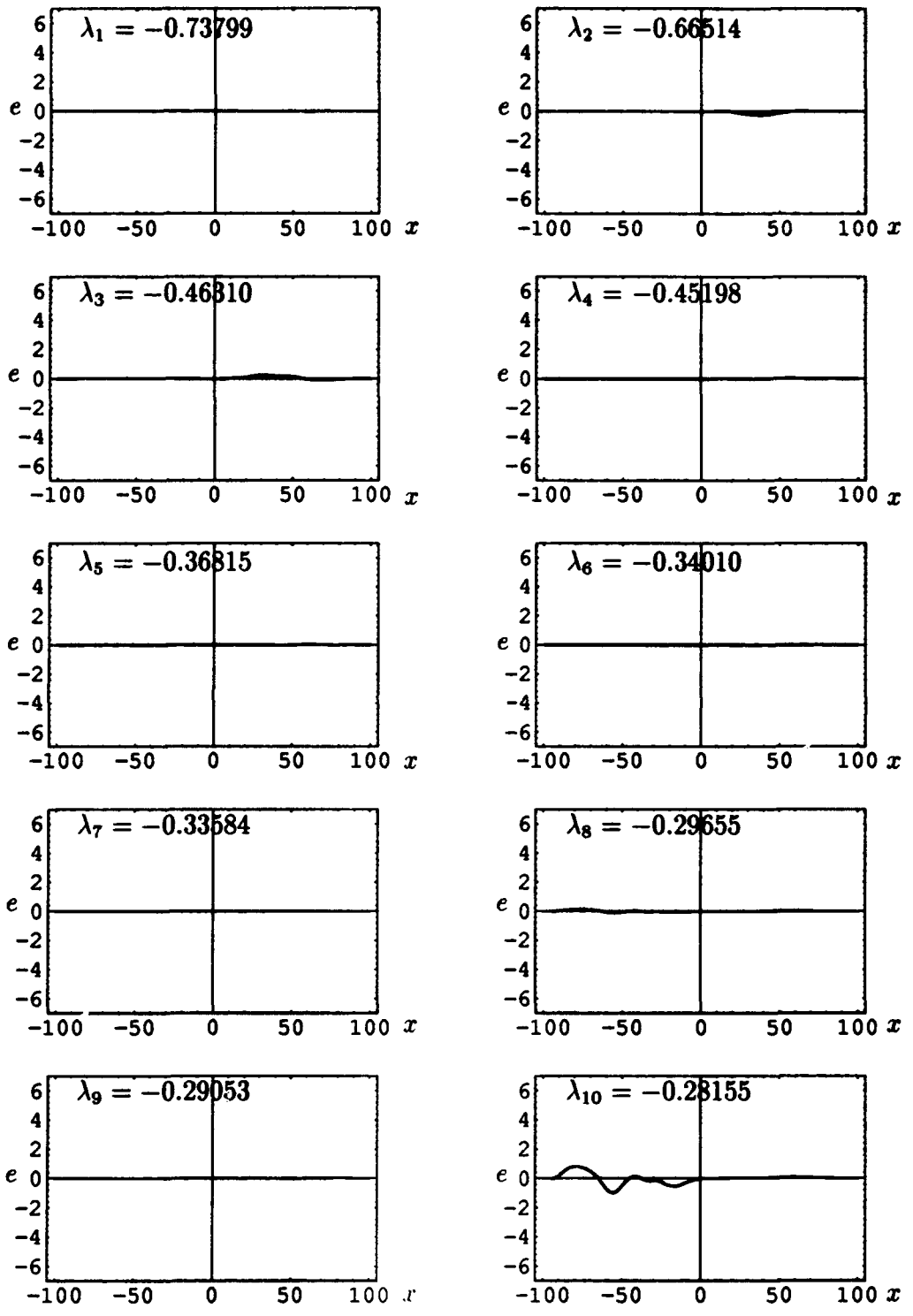


Figure A.56 Cross section at $y = 0$ of bending modes $1 - 10$.

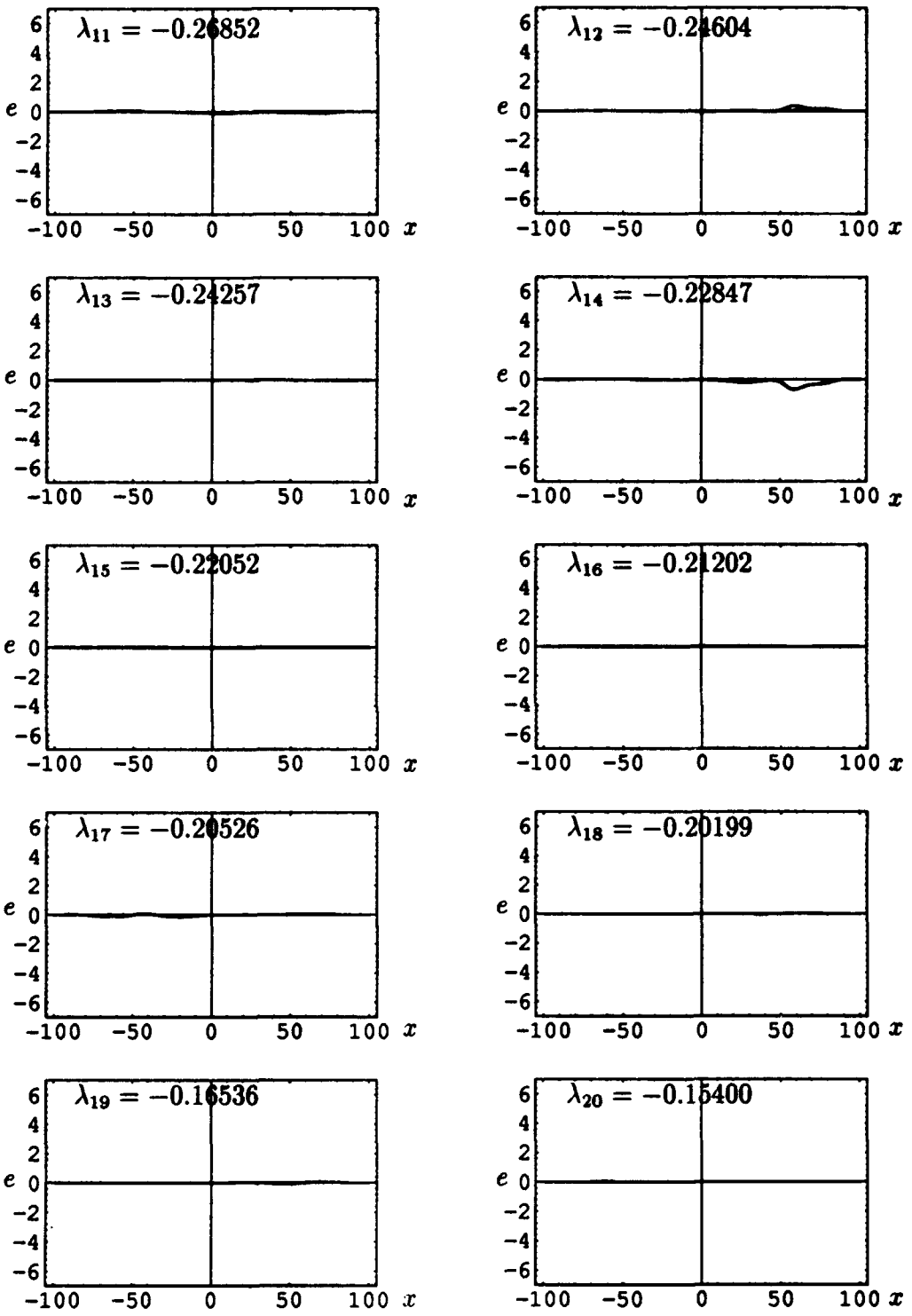


Figure A.57 Cross section at $y = 0$ of bending modes $11 - 20$.

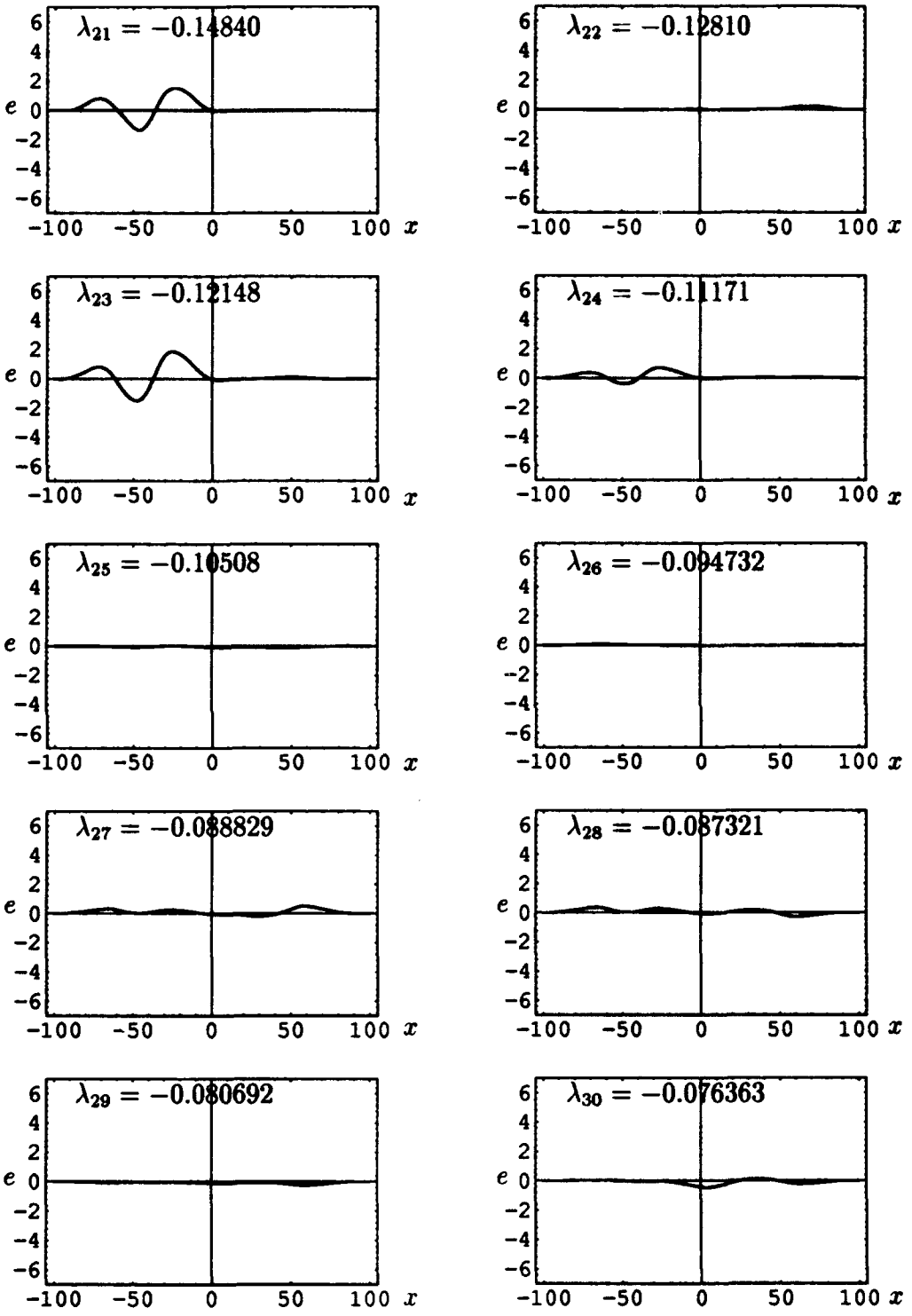


Figure A.58 Cross section at $y = 0$ of bending modes $21 - 30$.

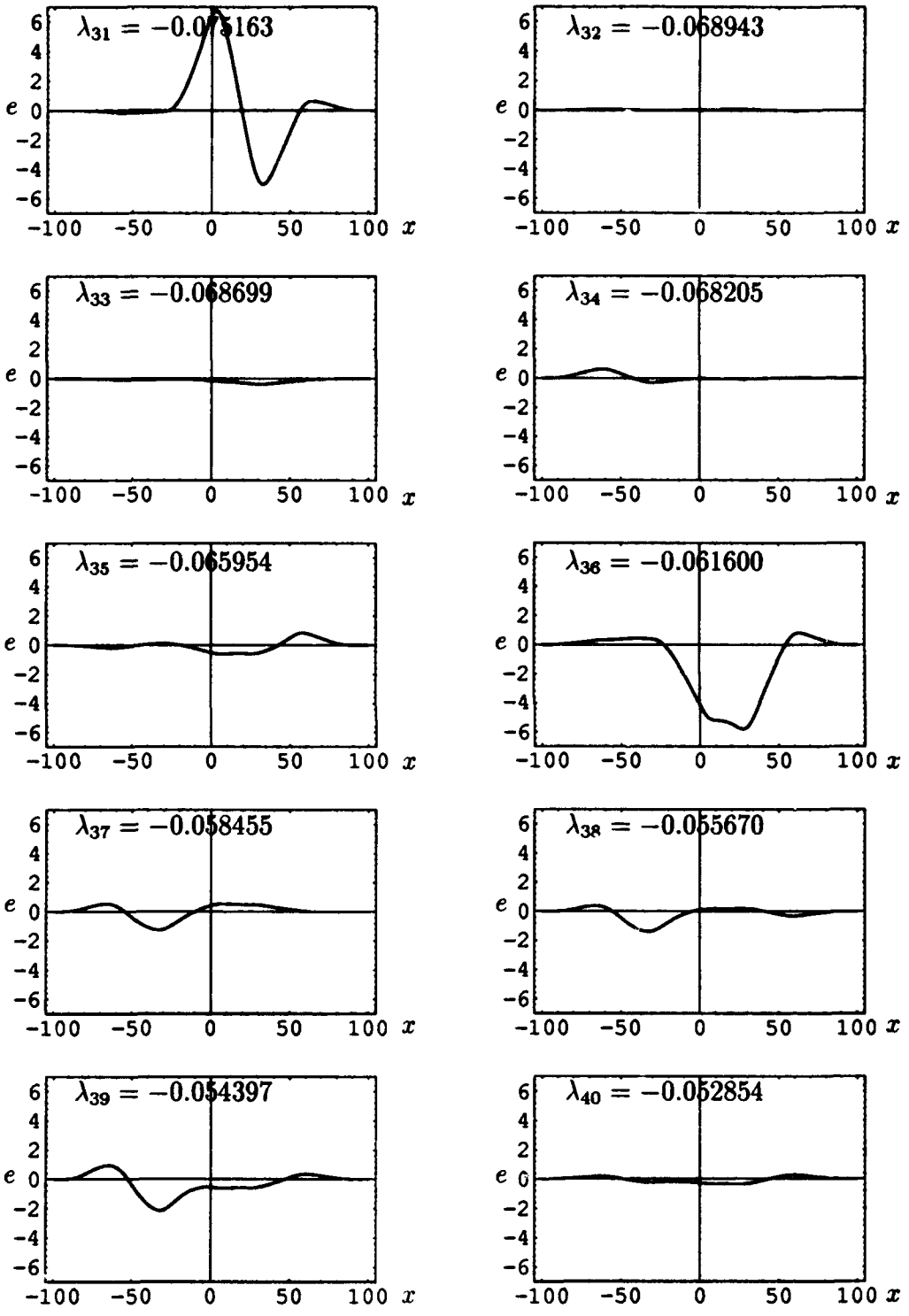


Figure A.59 Cross section at $y = 0$ of bending modes $31 - 40$.

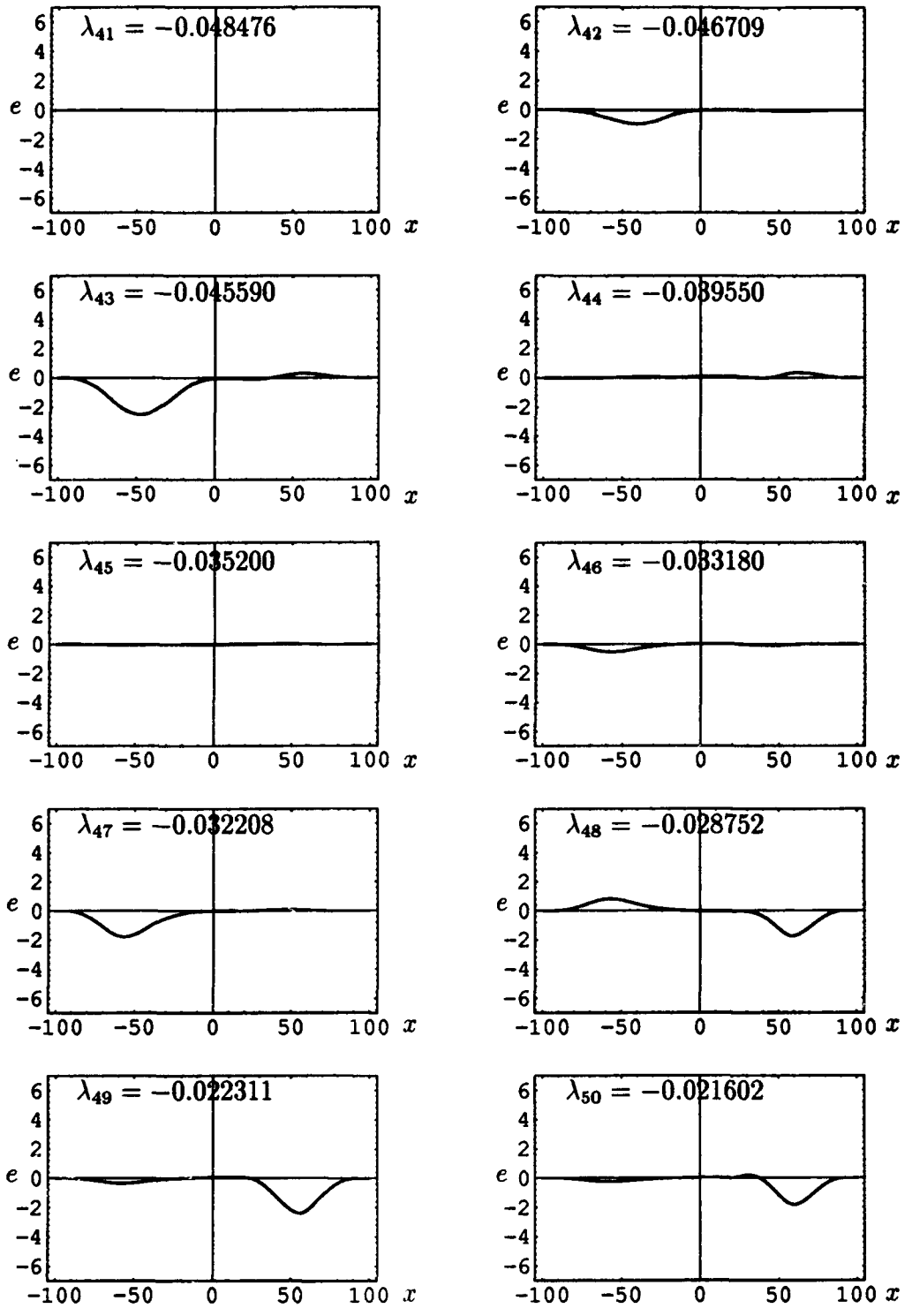


Figure A.60 Cross section at $y = 0$ of bending modes $41 - 50$.

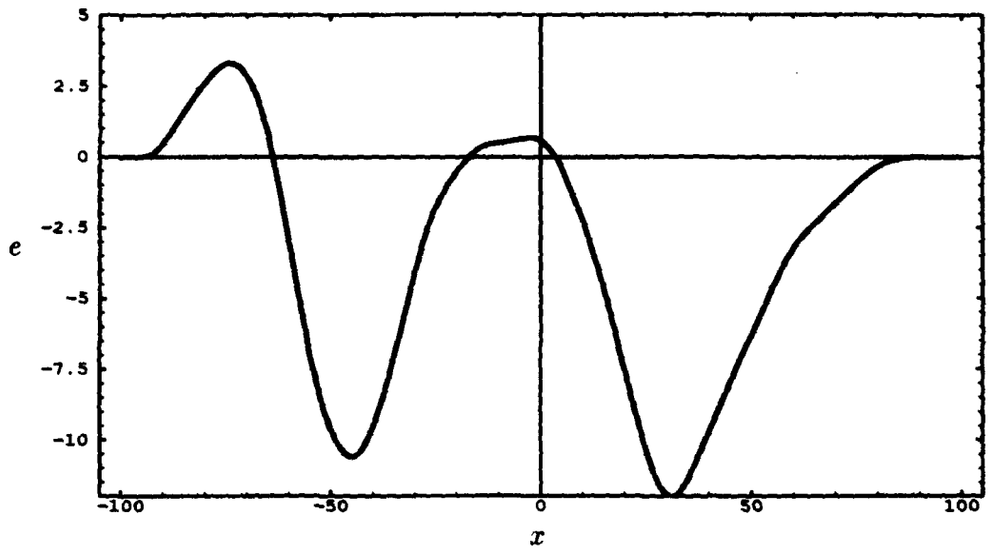


Figure A.61 Cross Section at $y = 0$: Sum of all fifty bending modes.

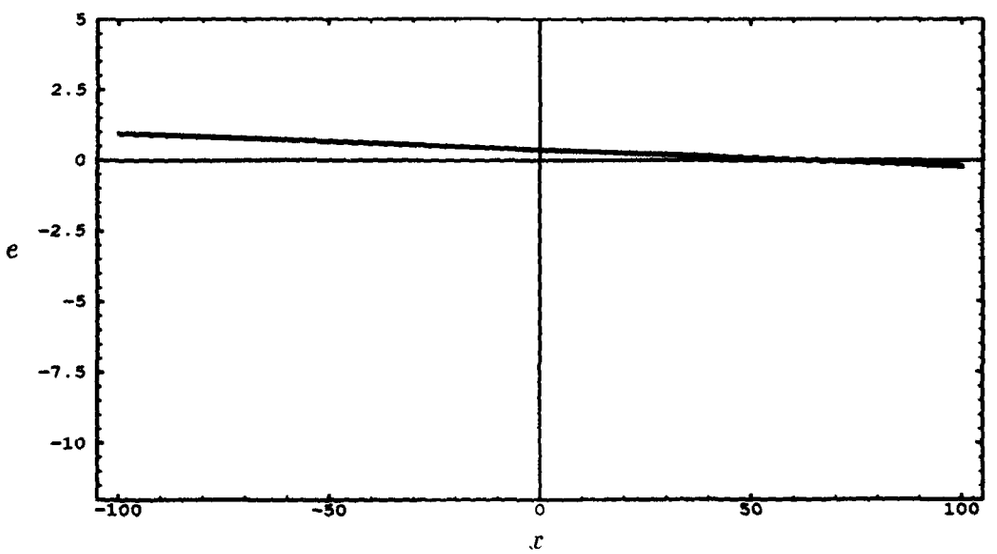


Figure A.62 Cross Section at $y = 0$: Regression plane.

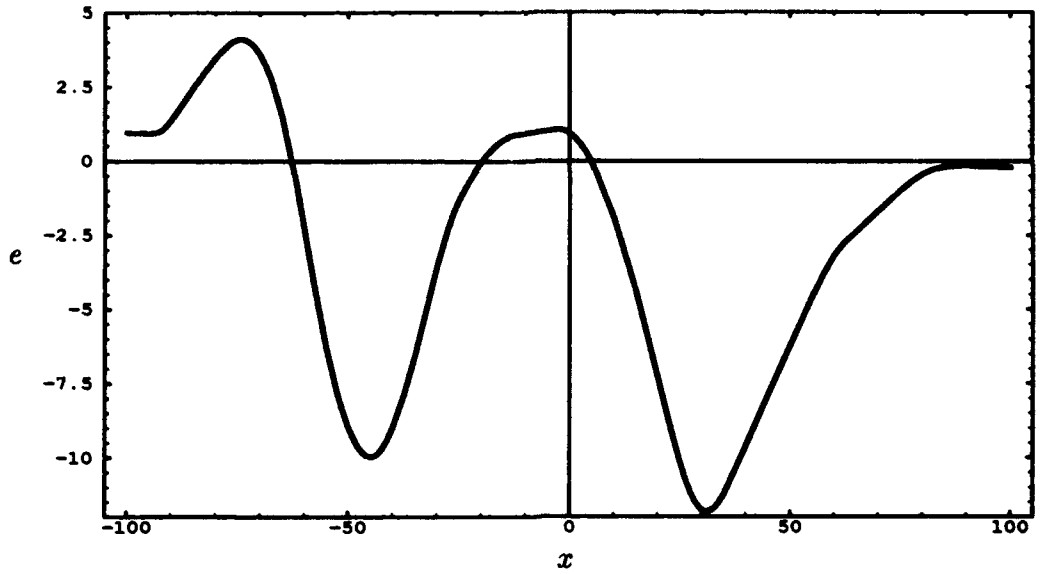


Figure A.63 Cross Section at $y = 0$: Sum of bending modes and regression plane.

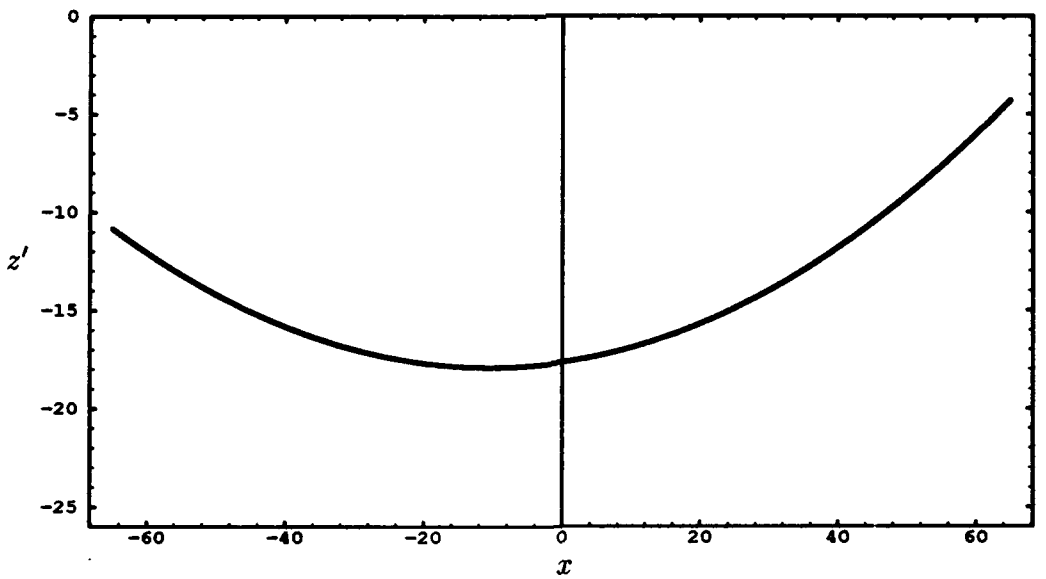


Figure A.64 Cross Section at $y = 0$: Polynomial trend (z') subtracted from physical z coordinates.

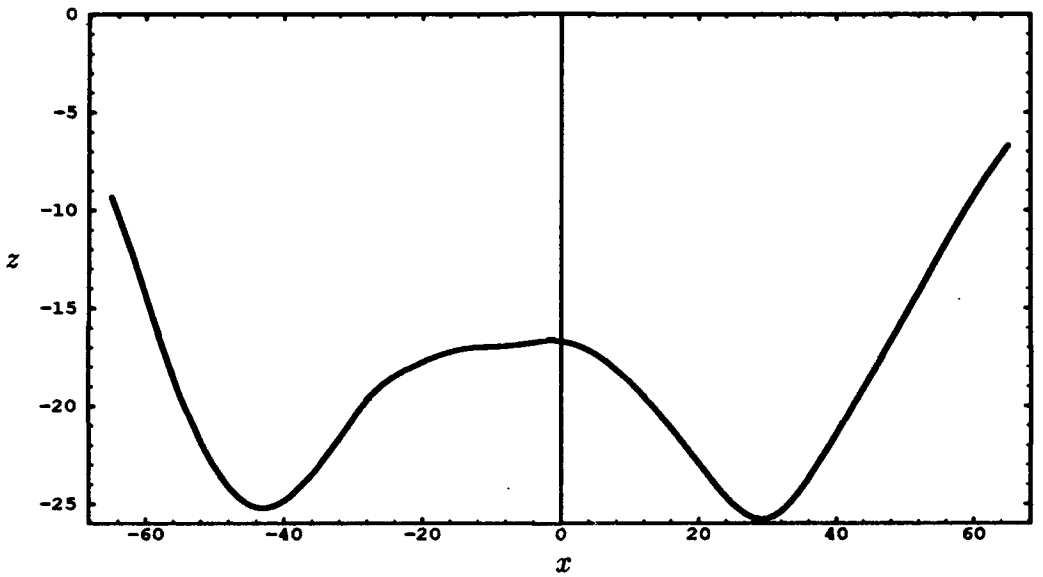


Figure A.65 Cross Section at $y = 0$: Estimated height of face in physical dimensions.

Appendix B. Eigenvalue Investigation

This appendix contains complete analysis for only three of the seal areas used to validate the trend discussed in Section 3.4.3: the Megg, Begg, and a sparsed seal area data file (plaster cast identification number 1918). The information presented here is similar to that presented for the original egg. The comparisons made in Section 3.4.3, and those presented in this appendix are only a few of all the investigations done for this research. All the investigations validate the premise that the eigenvalues of the energy matrix contain information which characterizes the seal area of a face. The eigenvalue trends show that smaller seal areas have larger bending energies compared to larger seal areas, and more protruding noses/chins have smaller bending energies. This appendix also includes the data files and two-dimensional plots of other egg modifications such as the Degg.

B.1 Megg

The Megg is a slight modification of the original egg described in Section 3.4.3; it has an additional two points representing the corners of the mouth. The Megg coordinate data is listed in Figure B.1 and plotted in Figure B.1. The following are modifications to the Megg and the identifying name:

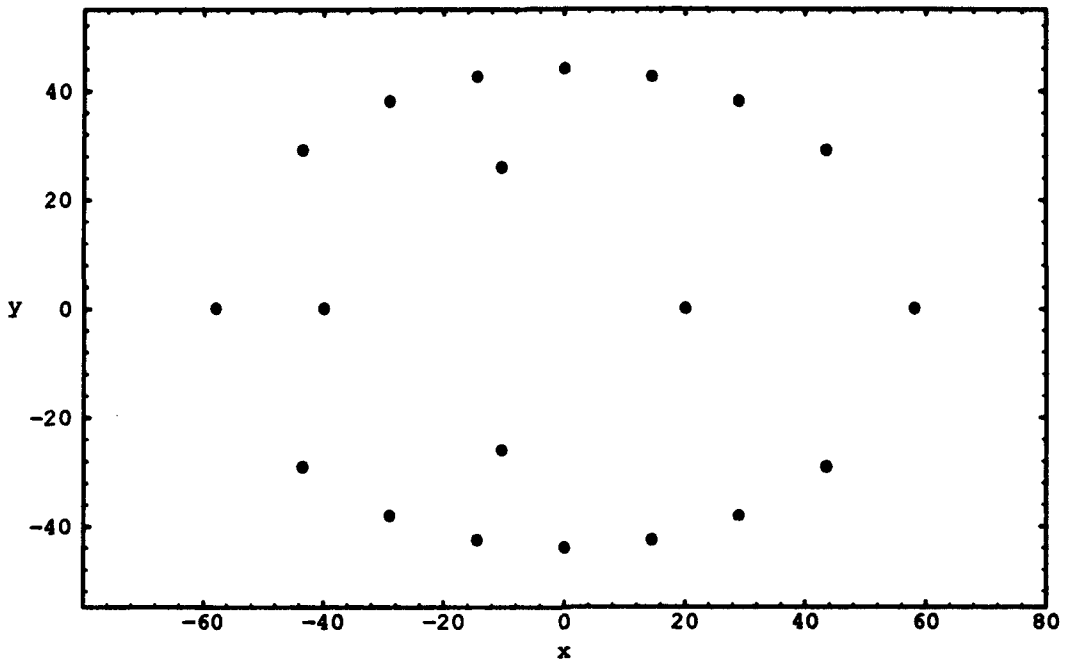
- The Megg is lengthed 30 % by scaling the x coordinates. (Megg-x130)
- The Megg is widened 30 % by scaling the y coordinates. (Megg-y130)
- The Megg is increased 30 % in both the x and y directions. (Megg-xy130)
- The Megg is increased 30 % in the x, y and z directions making it the same shape, but a different size. (Megg-y130)
- The height of the nose is changed by five units in both the positive and negative z direction creating two modified eggs. (Nose-up, Nose-down)

- The height of the chin is changed by five units in both the positive and negative z direction creating two modified eggs. (Chin-up, Chin-down)
- The position of the nose is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Nose-xpos, Nose-xneg)
- The position of the chin is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Chin-xpos, Chin-xneg)

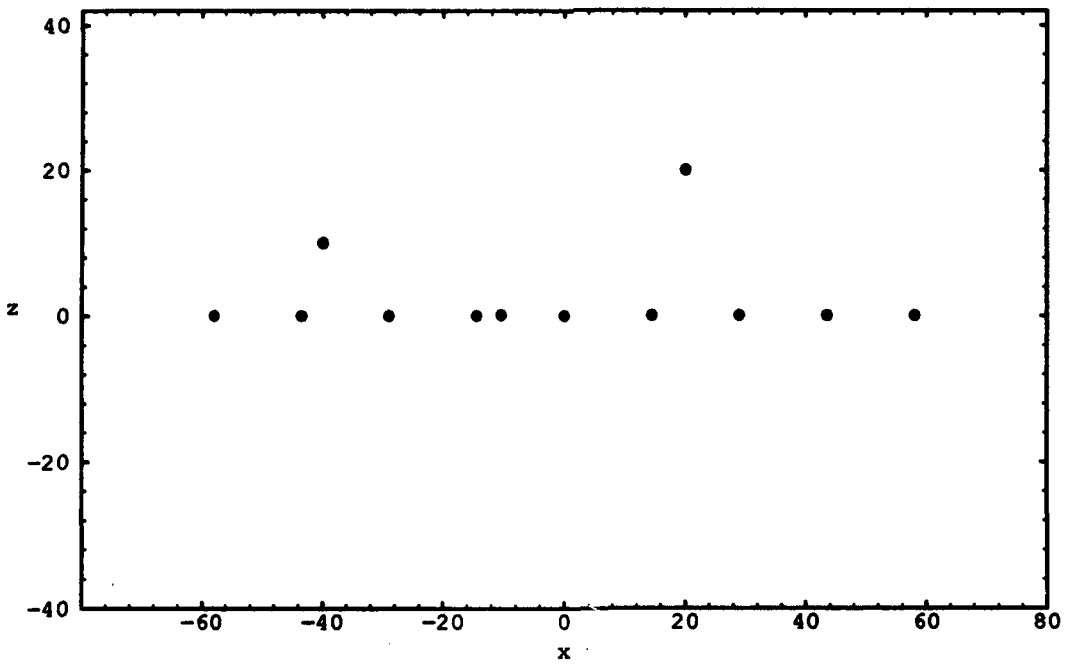
The eigenvalue plots for the various megg configurations (scaling, nose and chin height movement, and nose and chin position movement) are shown in Figures B.2 through B.6. In Figure B.2, the difference in eigenvalues between Megg-y130 and Megg-x130 is not as large as it is for Egg-x130 and Egg-y130 of the original egg in Section 3.4.3.

Elliptical Points	x	y	z
1	58.0	0.0000	0
2	43.5	29.1033	0
3	29.0	38.1051	0
4	14.5	42.6028	0
5	0.0	44.0000	0
6	-14.5	42.6028	0
7	-29.0	38.1051	0
8	-43.5	29.1033	0
9	-58.0	0.0000	0
10	-43.5	-29.1033	0
11	-29.0	-38.1051	0
12	-14.5	-42.6028	0
13	0.0	-44.0000	0
14	14.5	-42.6028	0
15	29.0	-38.1051	0
16	43.5	-29.1033	0
Nose	20.0	0.0000	20.0
Chin	-40.0	0.0000	10.0
19	-10.4	26.0000	0.1
20	-10.4	-26.0000	0.1

Table B.1 **Megg Coordinate Data:** Points 19 and 20 represent corners of the mouth.



(a) Megg Top View



(b) Megg Side View

Figure B.1 Megg Two-Dimensional Plots: The Megg which consists of 20 points; (a) Megg top view, (b) Megg side view.

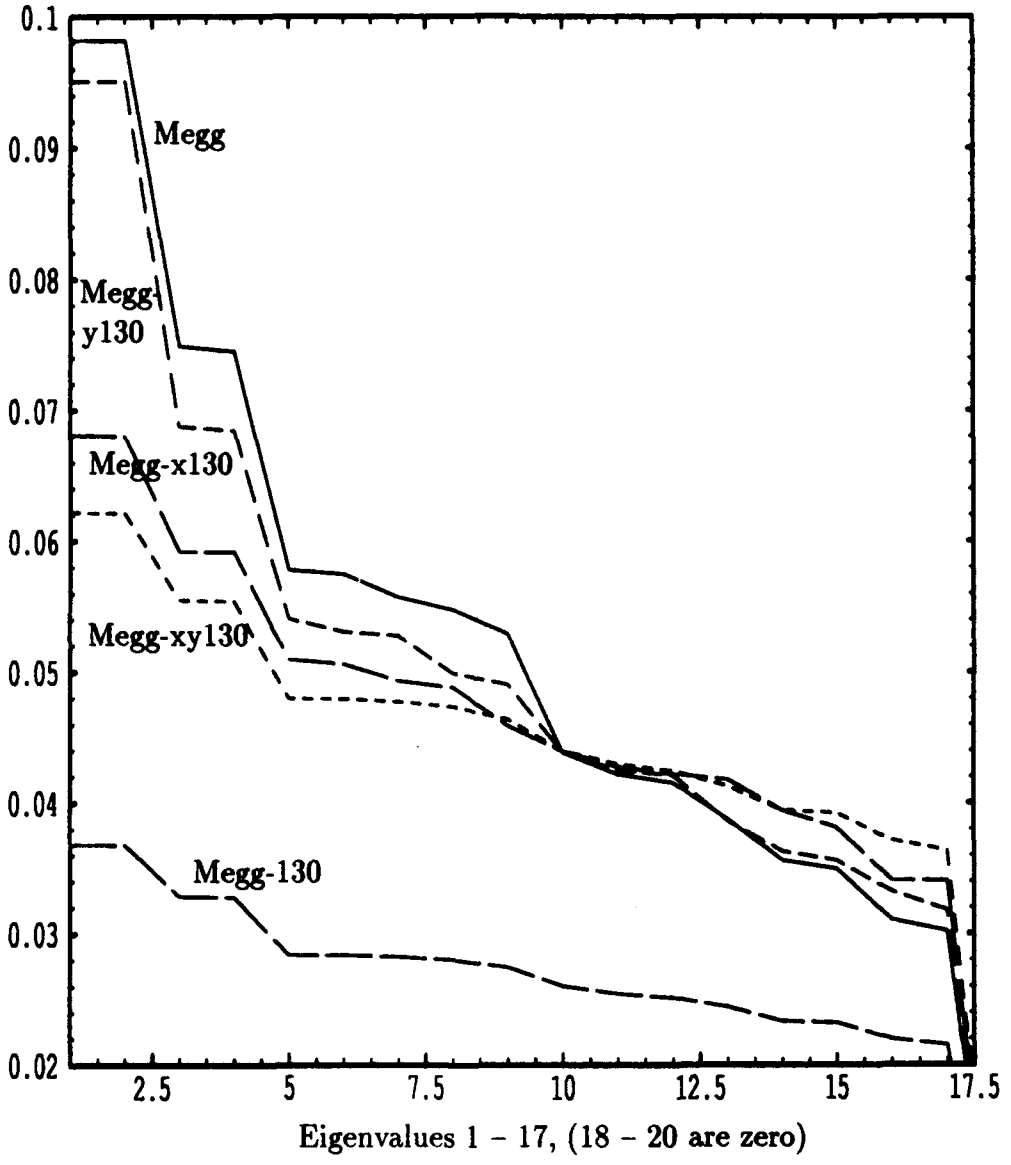


Figure B.2 Megg Scaling Effects Trend: Trend appearing when eigenvalues of the energy matrix from scaled megg data are compared. The vertical axis represents the magnitude of the eigenvalues.

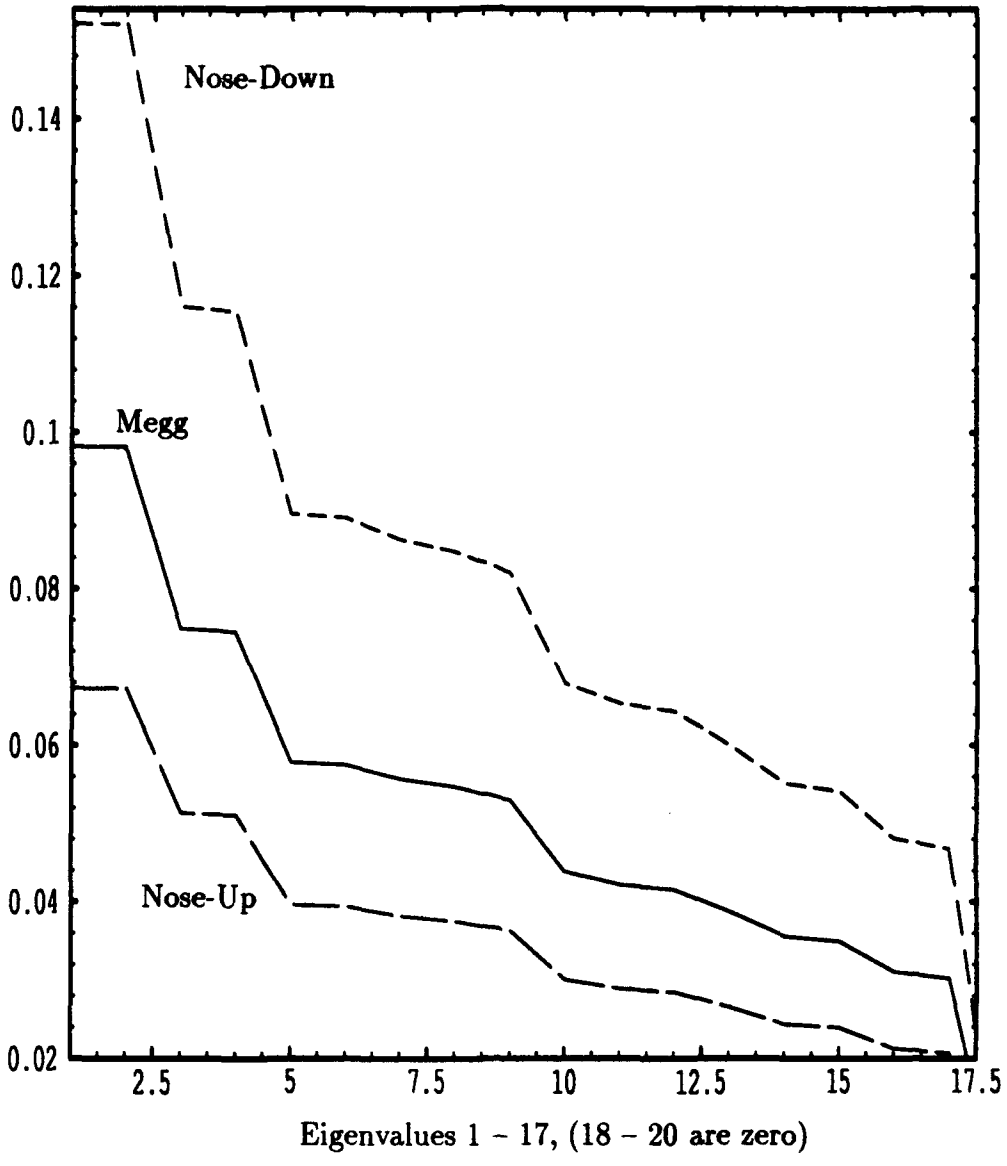


Figure B.3 Megg Nose Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

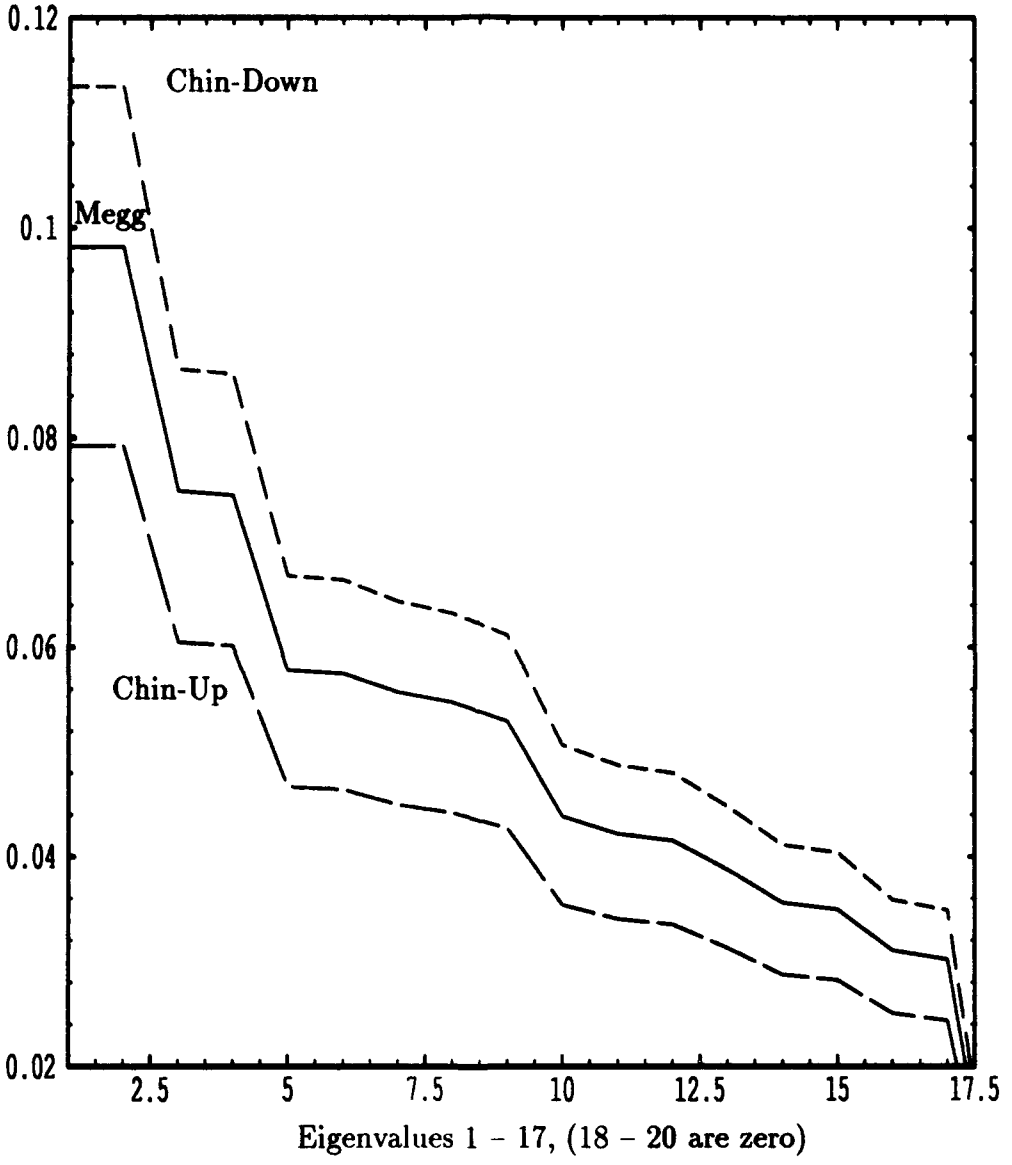


Figure B.4 Megg Chin Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

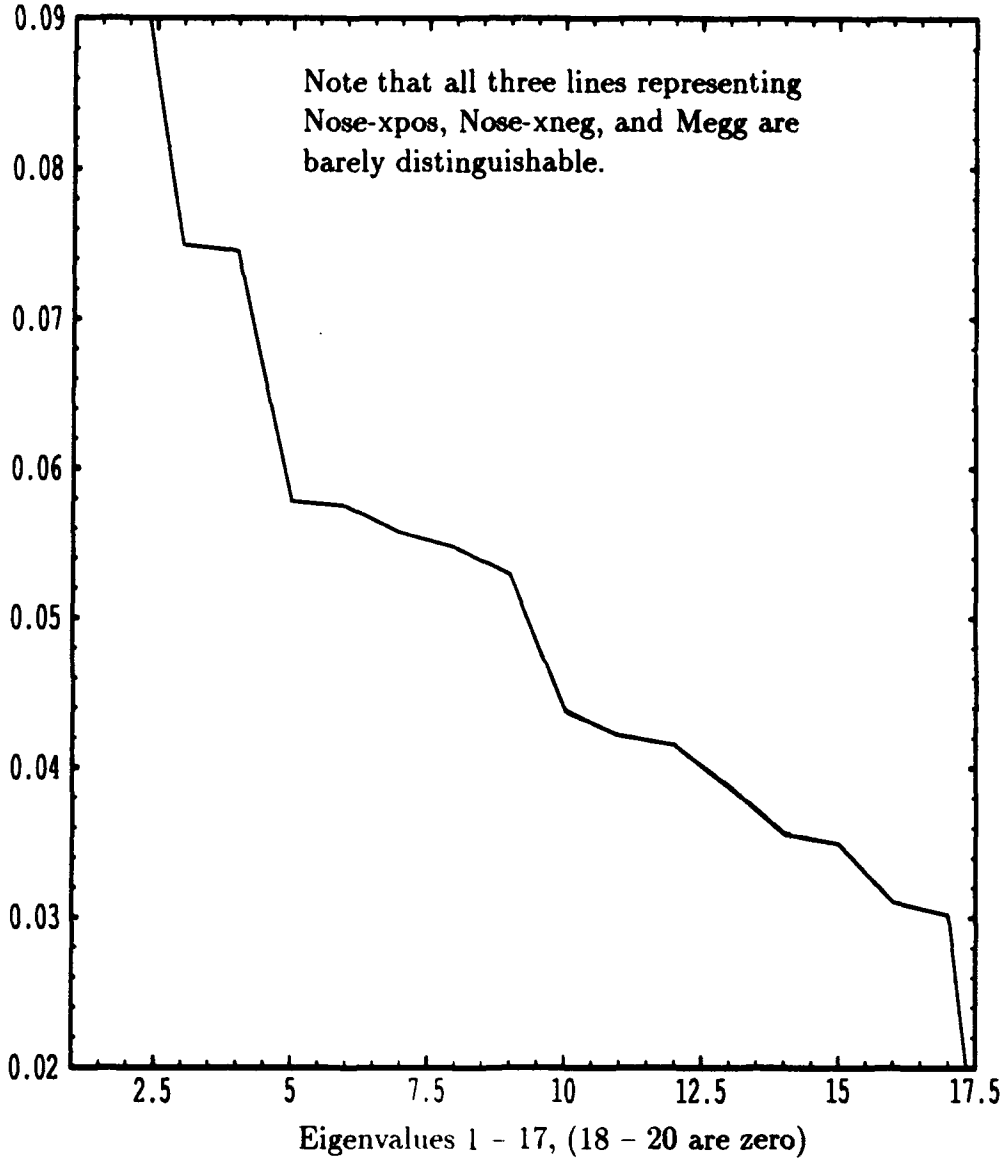


Figure B.5 Megg Nose Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

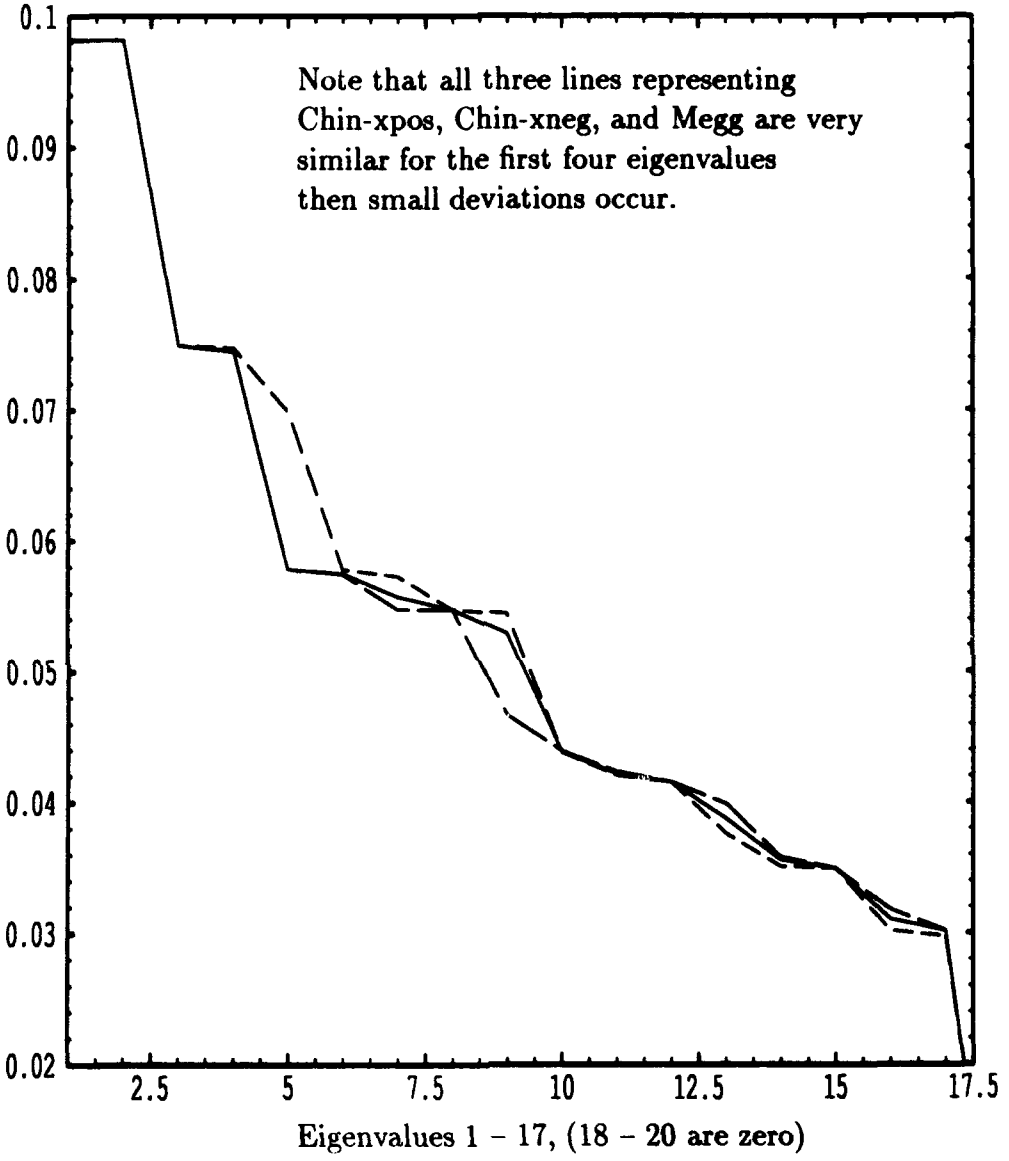


Figure B.6 Megg Chin Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

B.2 Begg

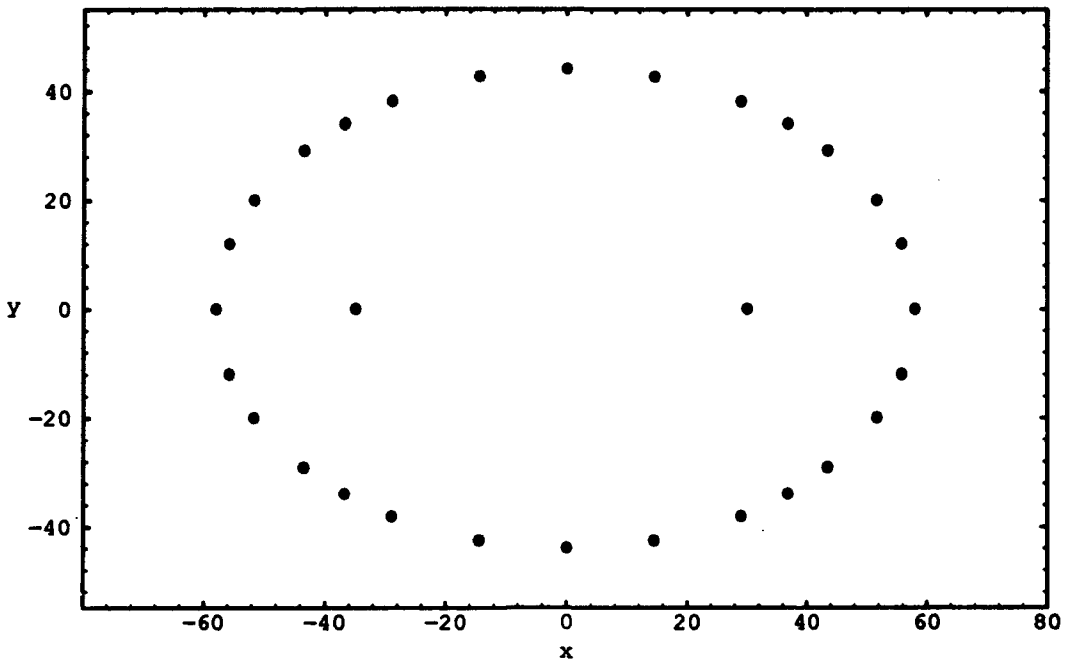
The Begg is a modification of the original egg discussed in Section 3.4.3; it has additional points located on the ellipse. The Begg coordinate data used is listed in Figure B.2 and plotted in Figure B.7. The following are modifications to the Begg and the identifying name:

- The Begg is lengthed 30 % by scaling the x coordinates. (Begg-x130)
- The Begg is widened 30 % by scaling the y coordinates. (Begg-y130)
- The Begg is increased 30 % in both the x and y directions.(Begg-xy130)
- The Begg is increased 30 % in the x, y and z directions making it the same shape, but a different size. (Begg-y130)
- The height of the nose is changed by five units in both the positive and negative z direction creating two modified eggs. (Nose-up, Nose-down)
- The height of the chin is changed by five units in both the positive and negative z direction creating two modified eggs. (Chin-up, Chin-down)
- The position of the nose is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Nose-xpos, Nose-xneg)
- The position of the chin is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Chin-xpos, Chin-xneg)

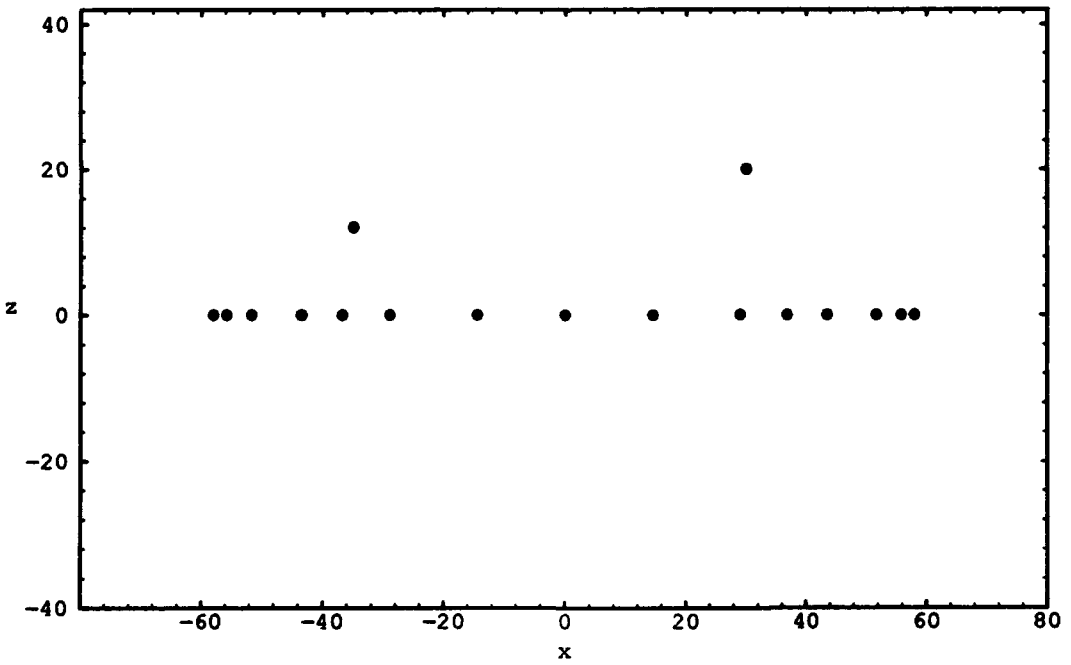
The eigenvalue plots for various begg configurations (scaling, nose and chin hieght movement, and nose and chin position movement) are shown in Figures B.8 through B.12. In Figure B.8, the difference in eigenvalues between Begg-y130 and Begg-x130 has narrowed even more than what it is for the Megg.

Elliptical Points	x	y	z
1	58.0	0.0000	0
2	55.8	12.0000	0
3	51.7	20.0000	0
4	43.5	29.1033	0
5	36.8	34.0000	0
6	29.0	38.1051	0
7	14.5	42.6028	0
8	0.0	44.0000	0
9	-14.5	42.6028	0
10	-29.0	38.1051	0
11	-36.8	34.0000	0
12	-43.5	29.1033	0
13	-51.7	20.0000	0
14	-55.8	12.0000	0
15	-58.0	0.0000	0
16	-55.8	-12.0000	0
17	-51.7	-20.0000	0
18	-43.5	-29.1033	0
19	-36.8	-34.0000	0
20	-29.0	-38.1051	0
21	-14.5	-42.6028	0
22	0.0	-44.0000	0
23	14.5	-42.6028	0
24	29.0	-38.1051	0
25	36.8	-34.0000	0
26	43.5	-29.1033	0
27	51.7	-20.0000	0
28	55.8	-12.0000	0
Nose	30.0	0.0000	20.0
Chin	-35.0	0.0000	12.0

Table B.2 Begg Coordinate Data.



(a) Begg Top View



(b) Begg Side View

Figure B.7 Begg Two-Dimensional Plots: The Begg which consists of 30 points; (a) Begg top view, (b) Begg side view.

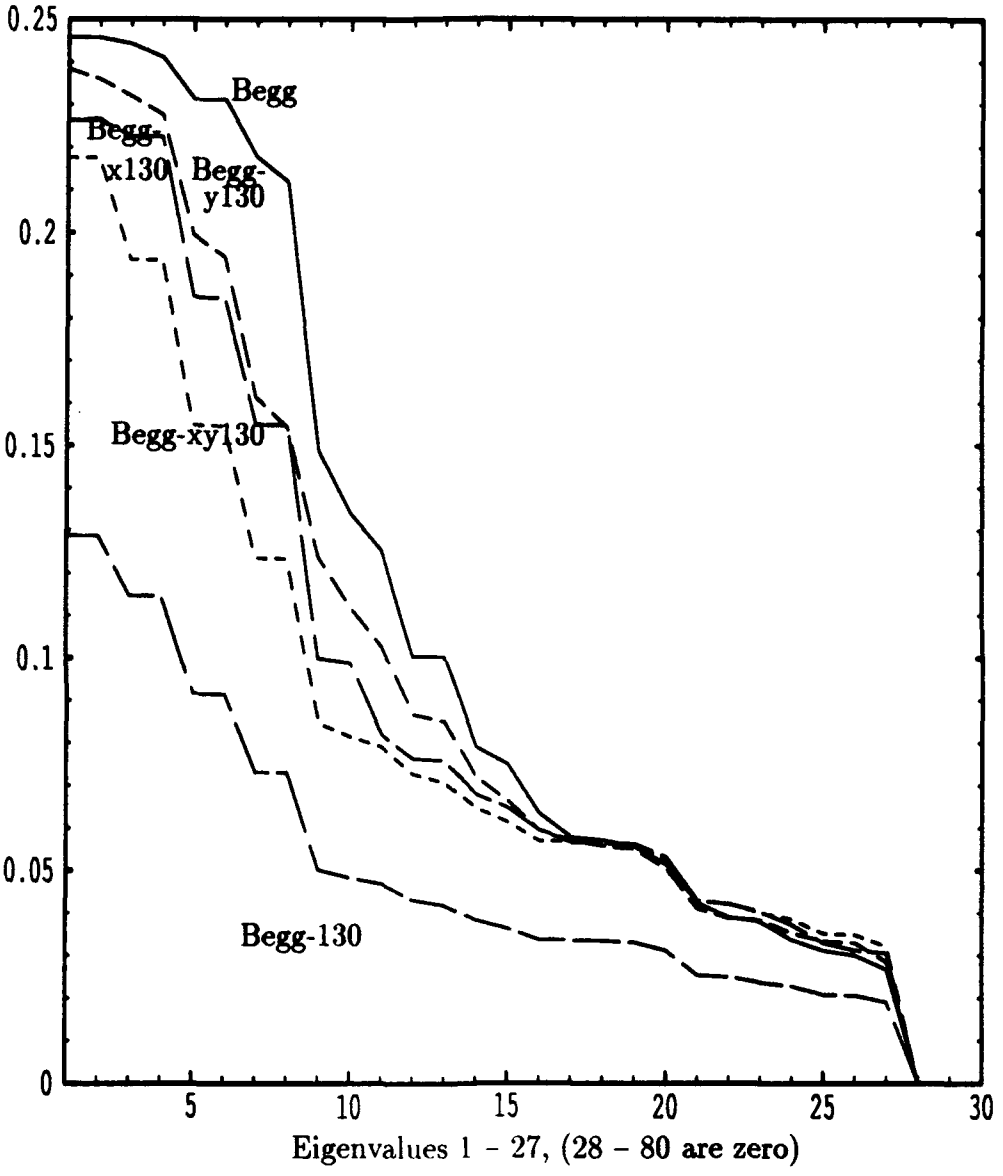


Figure B.8 **Begg Scaling Effects** Trend: Trend appearing when eigenvalues of the energy matrix from scaled begg data are compared. The vertical axis represents the magnitude of the eigenvalues.

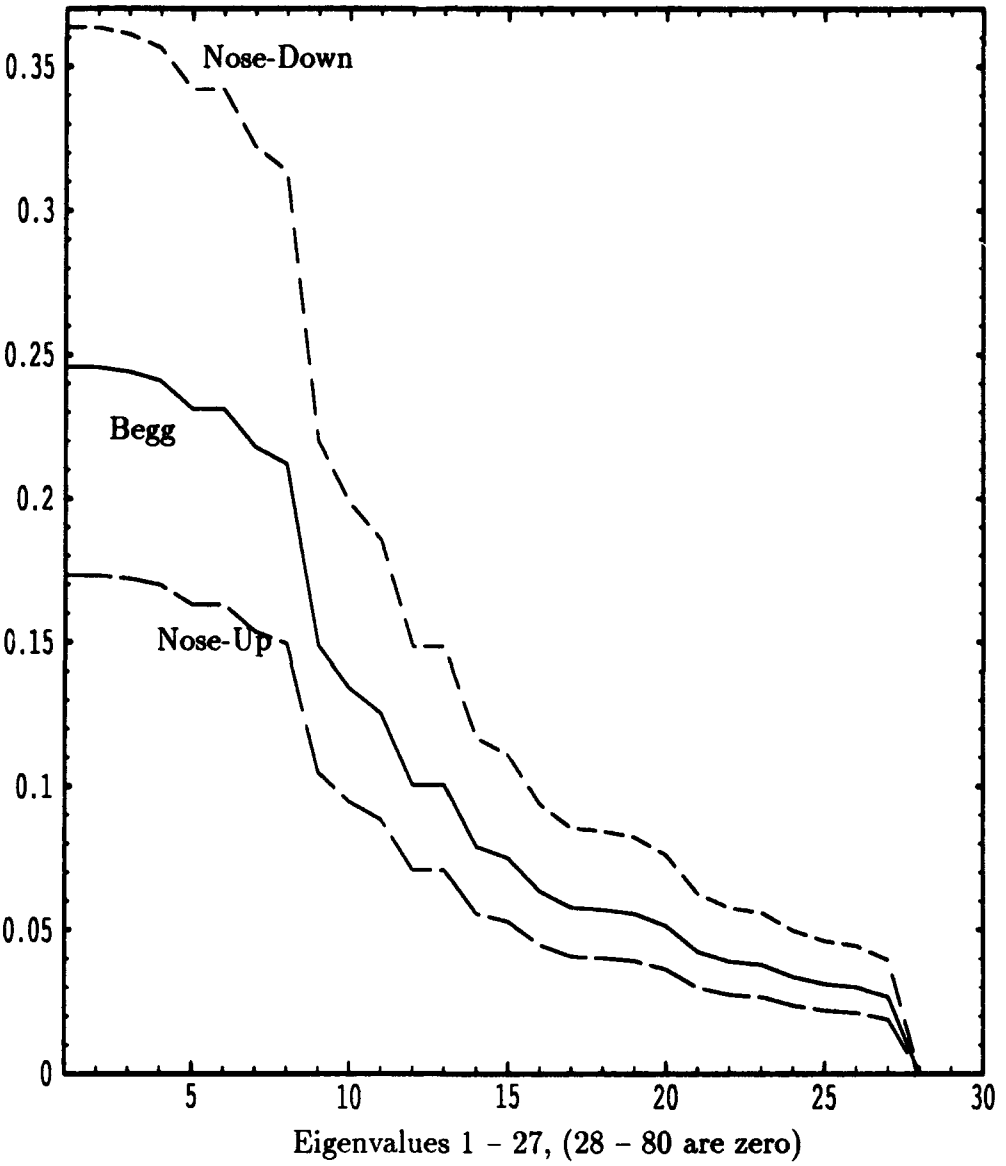


Figure B.9 Begg Nose Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

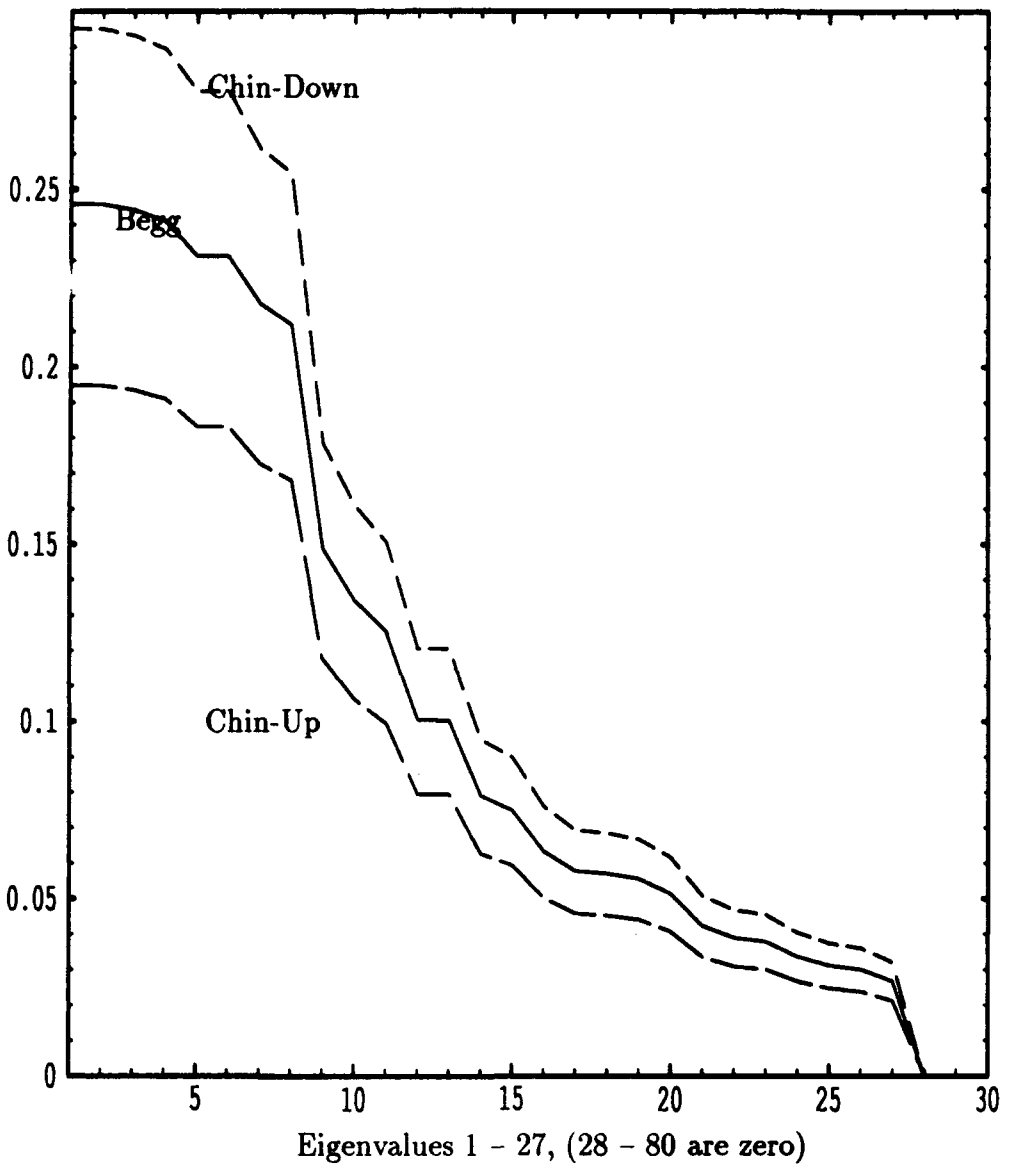


Figure B.10 Begg Chin Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

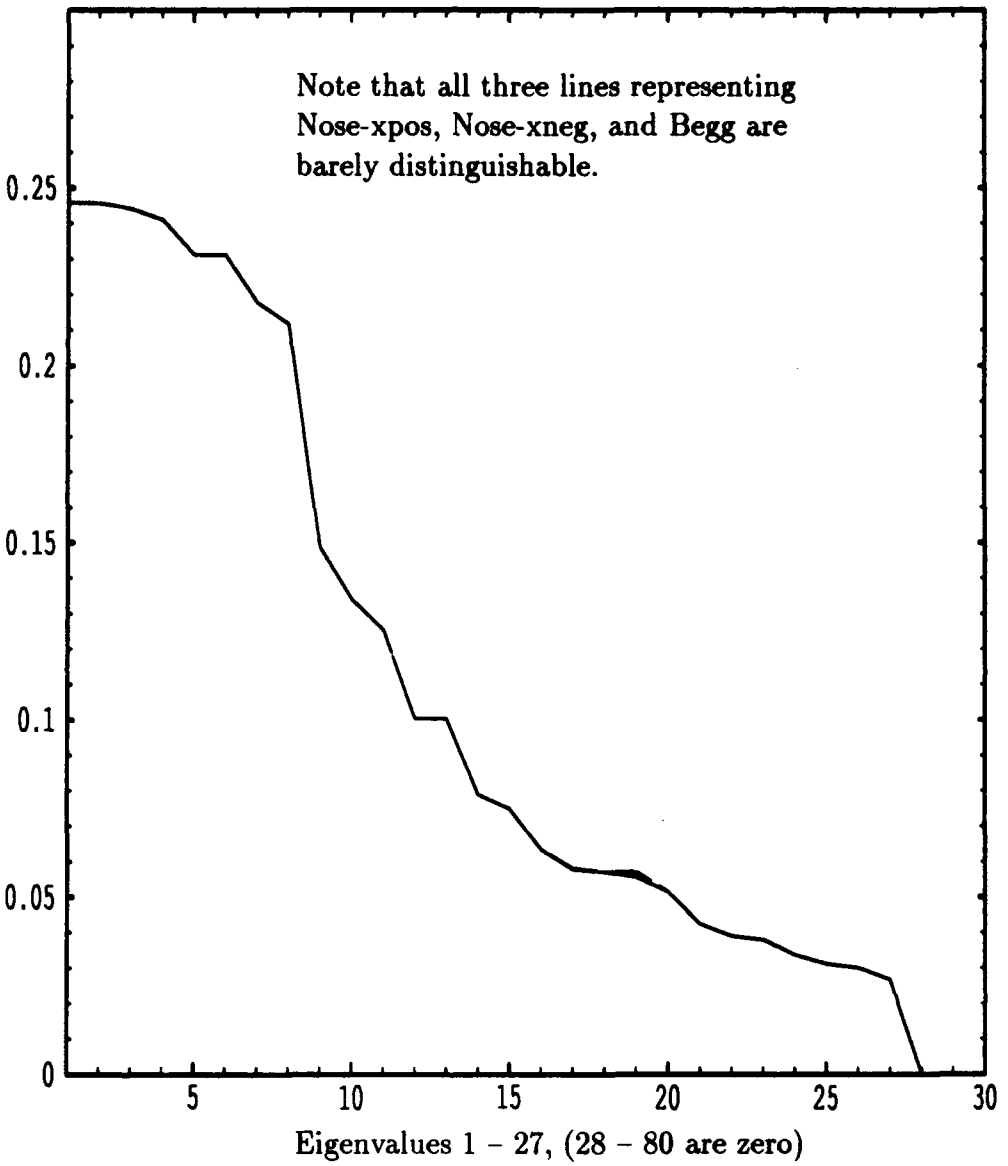


Figure B.11 Begg Nose Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

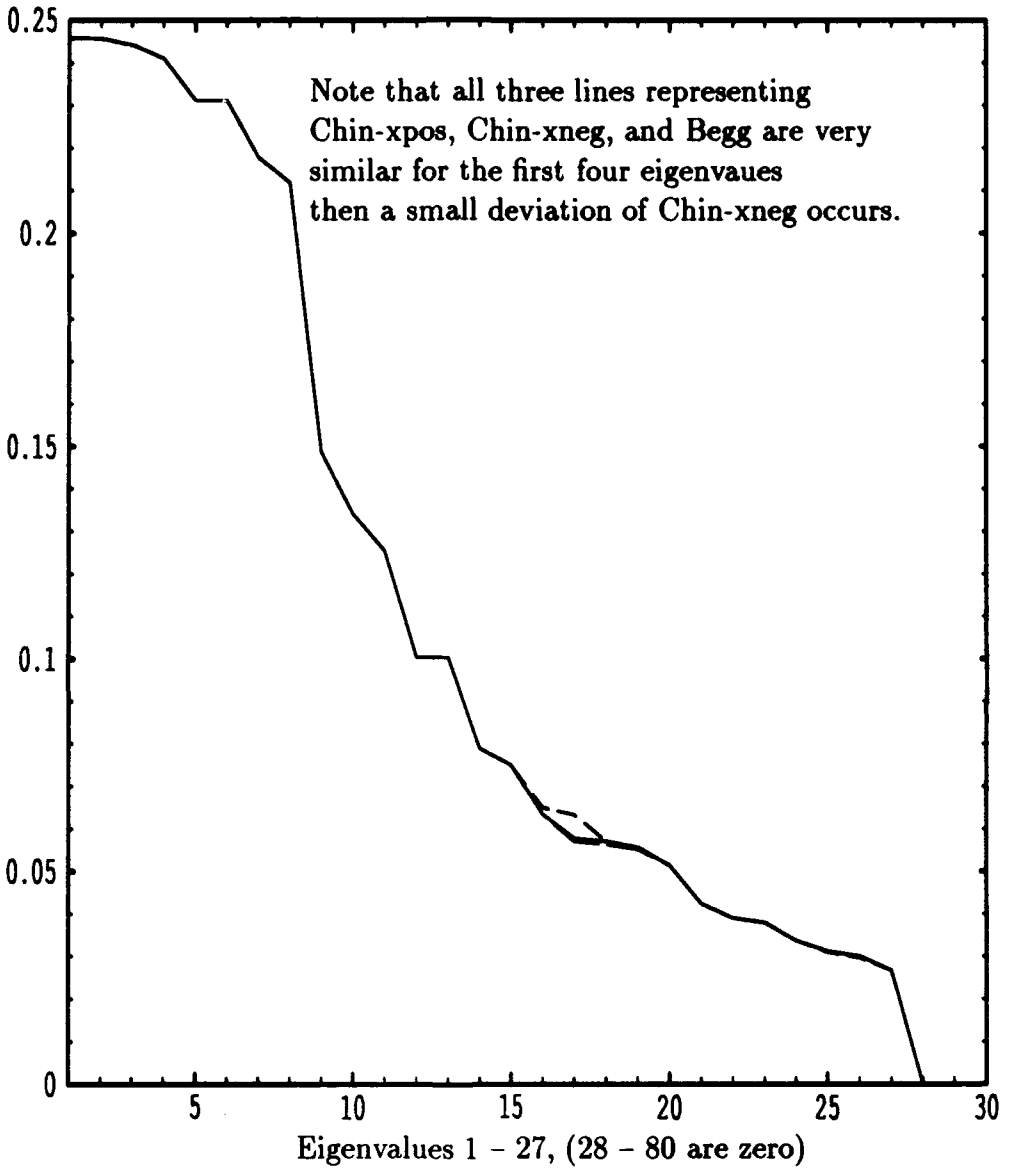


Figure B.12 Begg Chin Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

B.3 The Sparsed Seal Area

The sparsed seal area data presented contains 53 points, 39 along the seal line and 14 spacial delineators, refer to Section 5.1.2.3. The sparsed seal area coordinate data (Spar) used is listed in Table B.3 and plotted in Figure B.13. The following are modifications to the sparsed data and the identifying name:

- The sparsed data is lengthed 30 % by scaling the x coordinates. (Spar-x130)
- The sparsed data is widened 30 % by scaling the y coordinates. (Spar-y130)
- The sparsed data is increased 30 % in both the x and y directions. (Spar-xy130)
- The sparsed data is increased 30 % in the x, y and z directions making it the same shape, but a different size. (Spar-y130)
- The height of the nose is changed by five units in both the positive and negative z direction creating two modified eggs. (Nose-up, Nose-down)
- The height of the chin is changed by five units in both the positive and negative z direction creating two modified eggs. (Chin-up, Chin-down)
- The position of the nose is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Nose-xpos, Nose-xneg)
- The position of the chin is moved along the x-axis by five units in both the positive and negative x direction creating two modified eggs. (Chin-xpos, Chin-xneg)

The eigenvalue plots for the various sparsed seal area configurations (scaling, nose and chin hieght movements, and nose and chin position movements) are shown in Figures B.14 through B.18. In Figure B.14, the first eigenvalue of Spar-y130 is just slightly less than Spar, and the first eigenvalue of Spar-x130 is just slightly greater than Spar-xy130. After the fifth eigenvalue, both Spar-y130 and Spar-x130 are very similar. This is probably due to the increased number of points along the seal line.

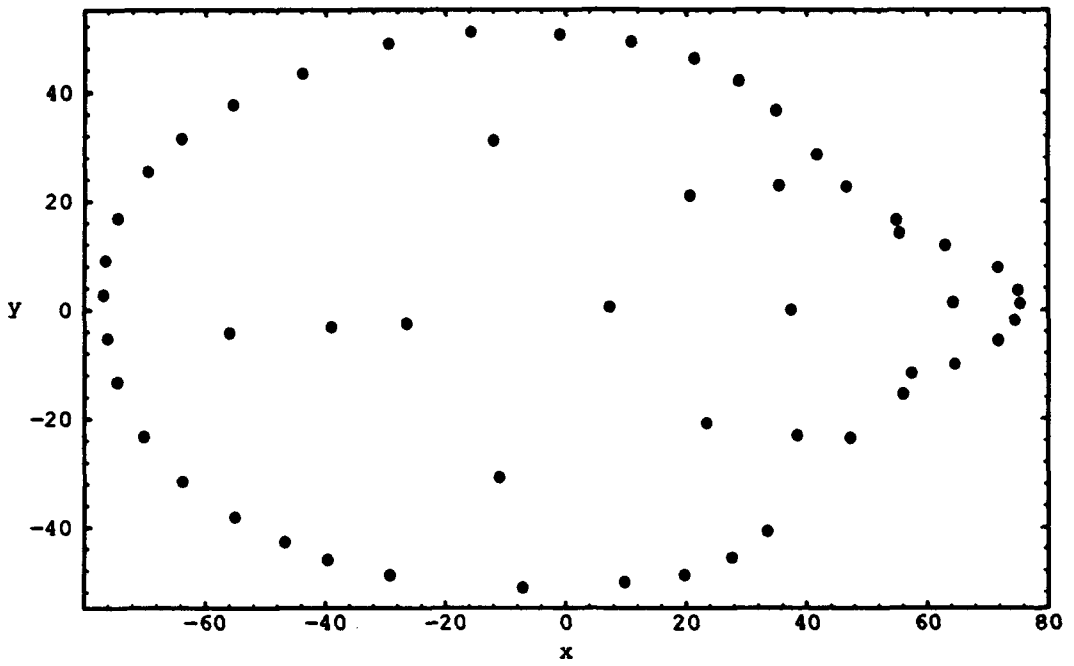
Elliptical Points	x	y	z
1	75.3535	1.0817	5.8181
2	75.0300	3.5596	5.5165
3	71.6694	7.7874	2.5061
4	62.9443	11.8356	-4.2388
5	54.8664	16.5134	-8.5468
6	46.5132	22.6889	-10.0646
7	41.6974	28.5294	-8.8402
8	34.9158	36.5712	-3.0690
9	28.6590	41.9883	1.3609
10	21.2669	46.0699	4.2047
11	10.8931	48.9706	4.3194
12	-1.06646	50.3352	2.4471
13	-15.9515	50.8339	1.6566
14	-29.5916	48.7260	1.8970
15	-43.9652	43.3279	0.2222
16	-55.4673	37.6029	-1.5748
17	-64.0308	31.3819	-2.8301
18	-69.5895	25.3499	-2.9655
19	-74.5105	16.7834	-2.8263
20	-76.5366	8.9360	-1.9456
21	-76.8219	2.6442	-1.4316
22	-76.1938	-5.3707	-1.1406
23	-74.5915	-13.3583	-1.4927
24	-70.1770	-23.3732	-0.3726
25	-63.8443	-31.6359	-0.0097
26	-55.1023	-38.2173	-0.8278
27	-46.7816	-42.8015	-0.6409
28	-39.6643	-46.1180	-0.6140
29	-29.3433	-48.9500	0.0766
30	-7.2097	-51.2553	-1.2595

Table B.3 Sparsed Seal Area Coordinate Data: First 30 points of sparsed seal line.

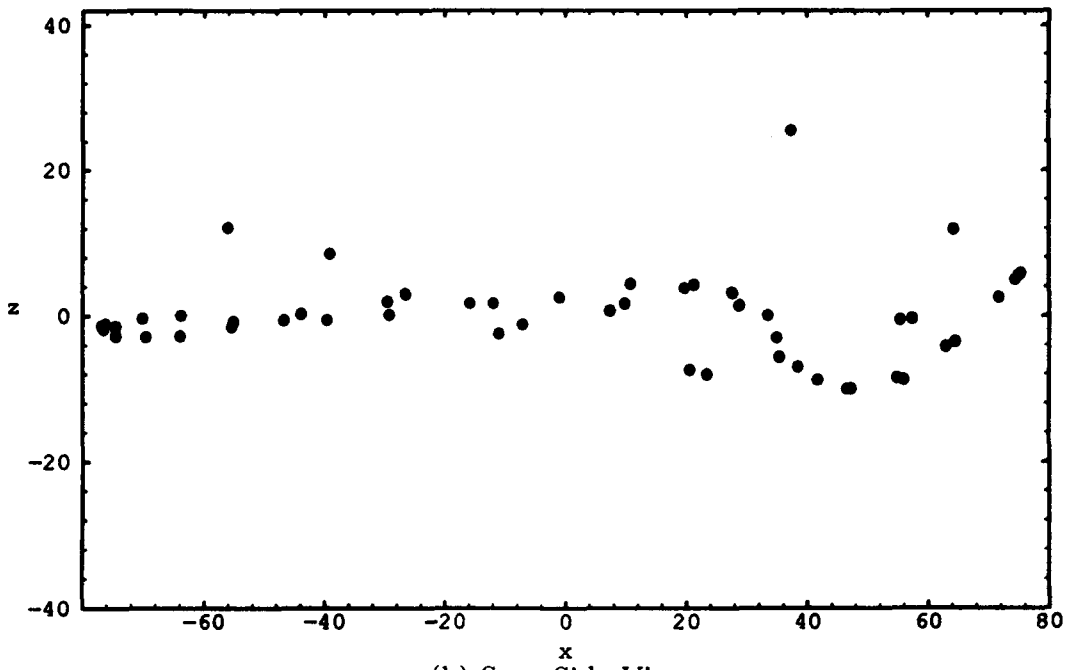
(continued on following page)

Elliptical Points	x	y	z
31	9.8409	-50.2378	1.6152
32	19.6829	-48.9064	3.7401
33	27.5739	-45.7295	3.0875
34	33.4359	-40.8089	-0.0001
35	47.2234	-23.5921	-10.047
36	55.9552	-15.5143	-8.8086
37	64.5642	-10.0478	-3.5586
38	71.6957	-5.6506	2.5156
39	74.4026	-1.9702	4.9482
40	57.3407	-11.7034	-0.4100
41	64.2647	1.26816	11.8555
42	55.3632	14.0791	-0.5817
43	38.4175	-23.0996	-7.0621
44	37.2901	-0.0846	25.4365
45	35.3161	22.8596	-5.7490
46	23.3749	-20.8377	-8.0985
47	7.3363	0.4003	0.6586
48	20.5686	21.0038	-7.4595
49	-11.0403	-30.8569	-2.5253
50	-11.9916	31.0462	1.6530
51	-26.5475	-2.6282	2.9290
52	-39.1494	-3.2615	8.4771
53	-56.0975	-4.3406	12.0504

Table B.3.b. Sparsed Seal Area Coordinate Data Continued: Points 40 – 53 are the 14 spacial delineators of sparsed seal area.



(a) Spar Top View



(b) Spar Side View

Figure B.13 Spar Two-Dimensional Plots: The Spar which consists of 53 points; (a) Spar top view, (b) Spar side view.

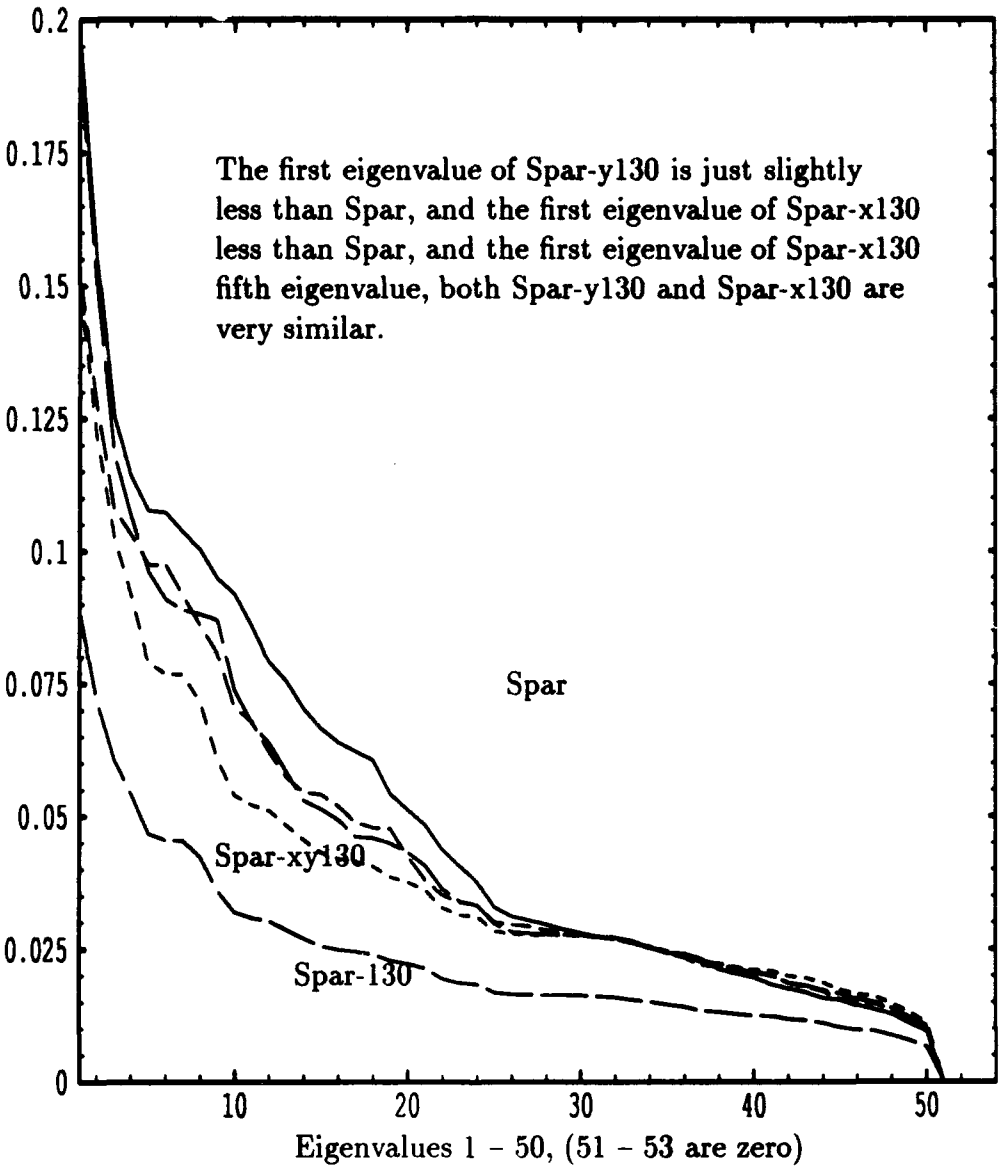


Figure B.14 Spar Scaling Effects Trend: Trend appearing when eigenvalues of the energy matrix from scaled spar data are compared. The vertical axis represents the magnitude of the eigenvalues.

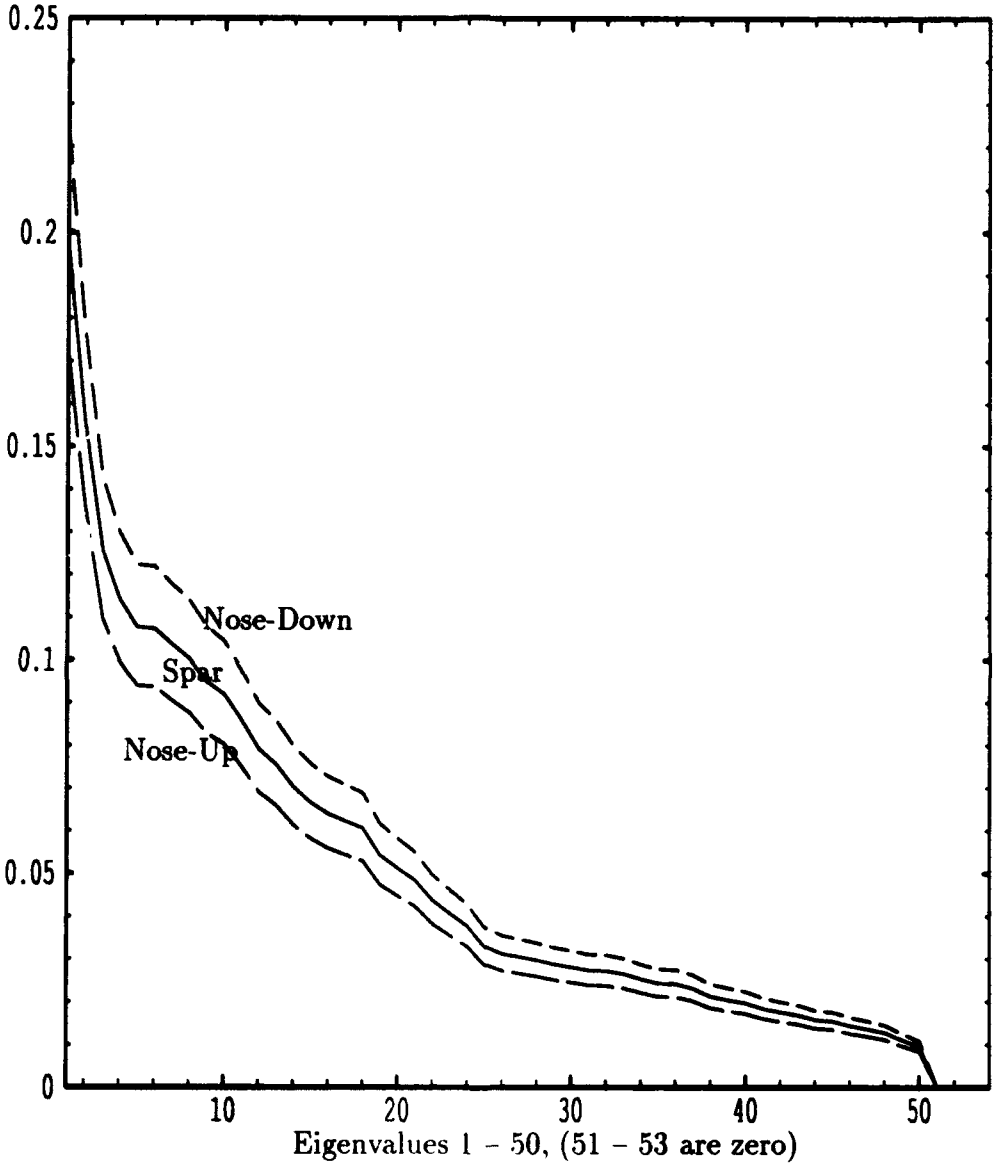


Figure B.15 Spar Nose Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

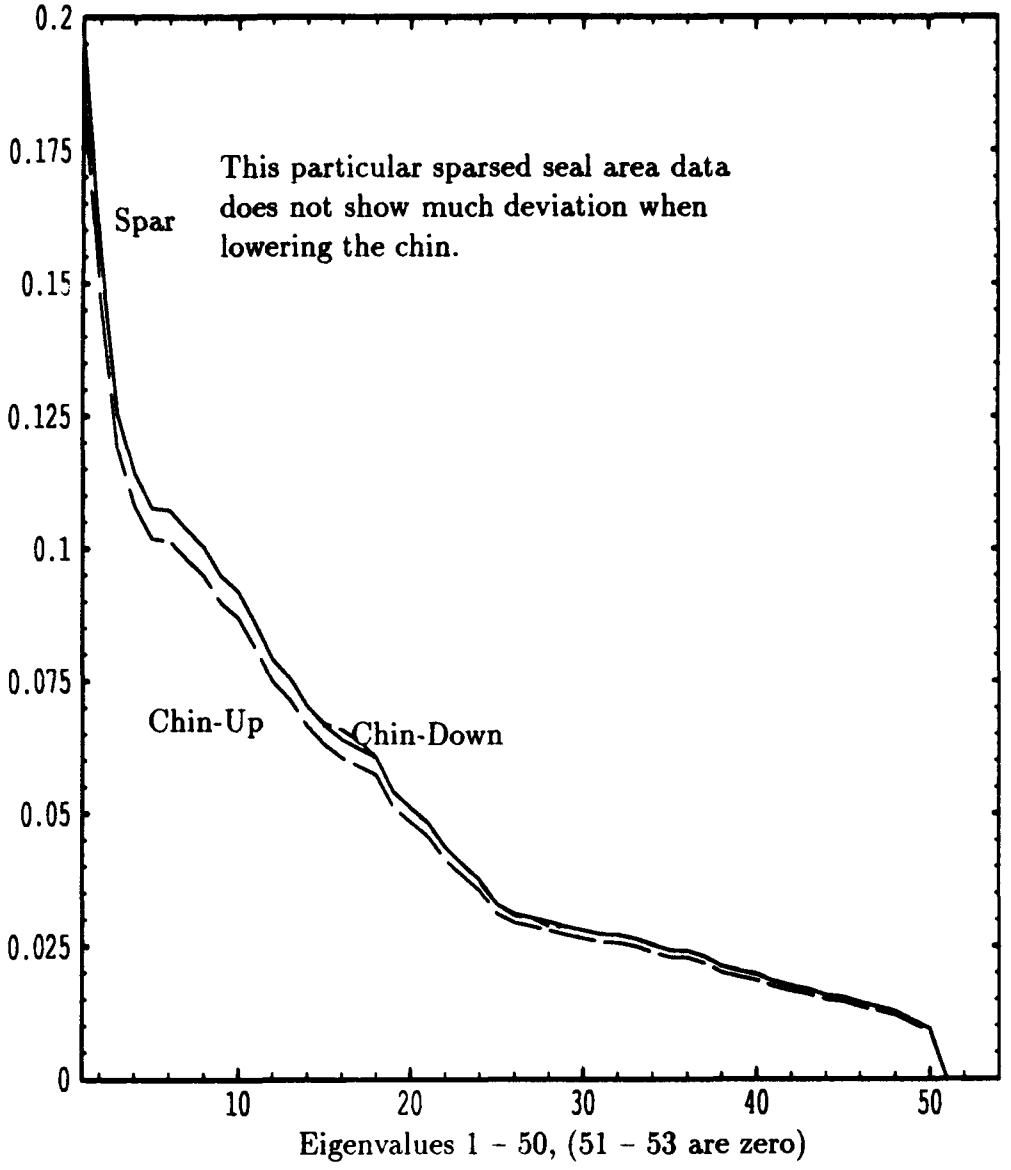


Figure B.16 Spar Chin Height Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in height of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

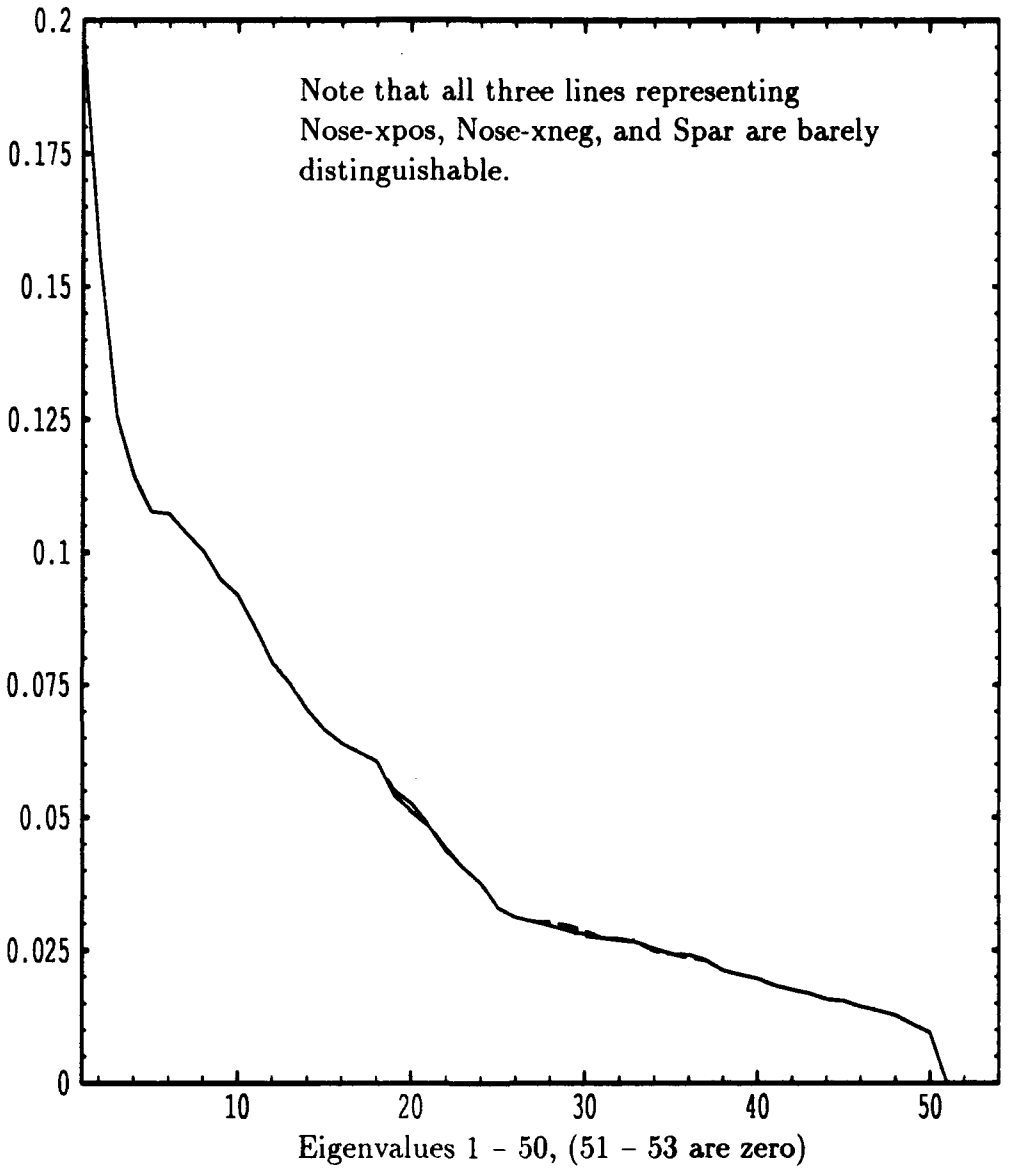


Figure B.17 Spar Nose Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the nose are compared. The vertical axis represents the magnitude of the eigenvalues.

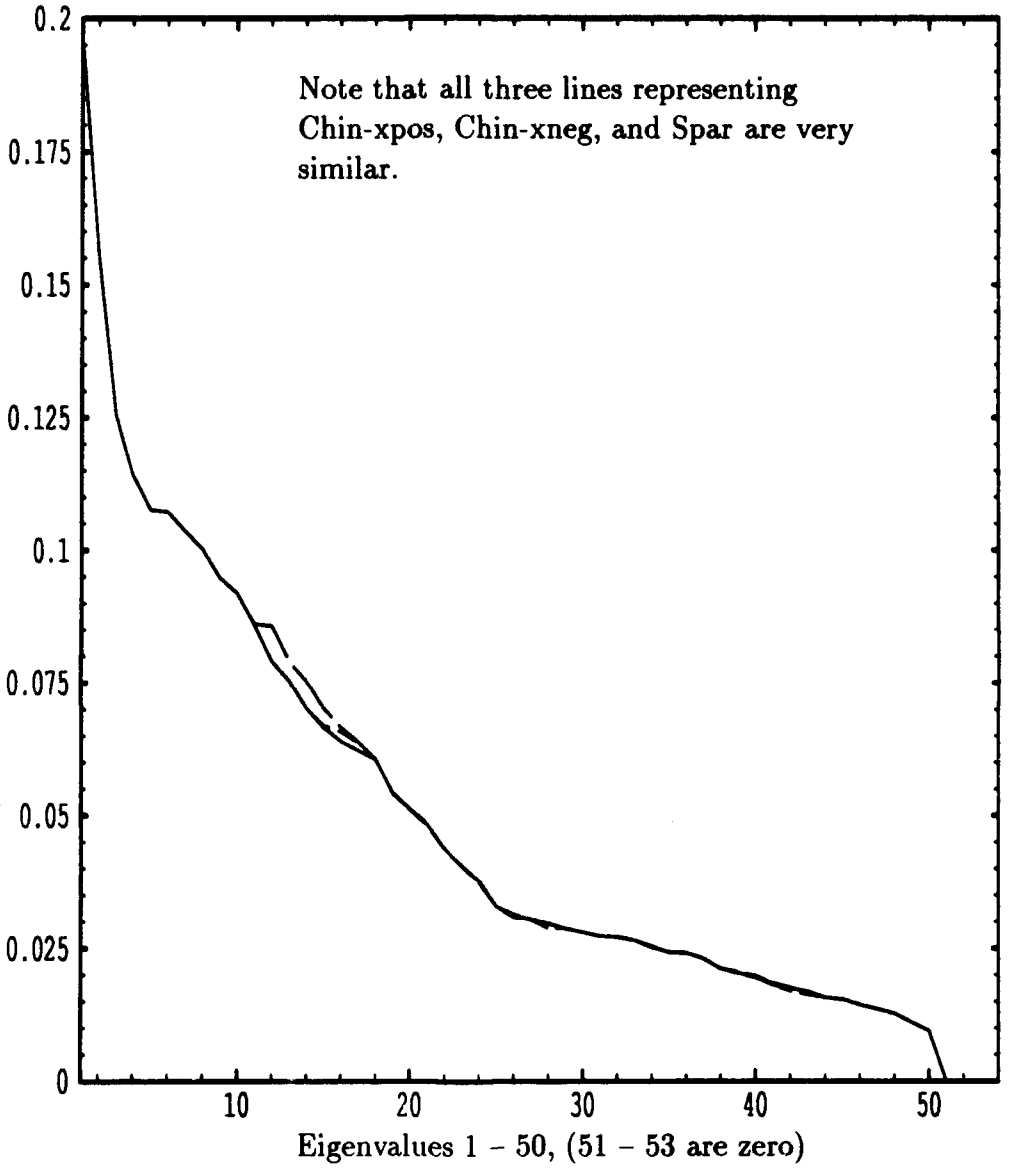
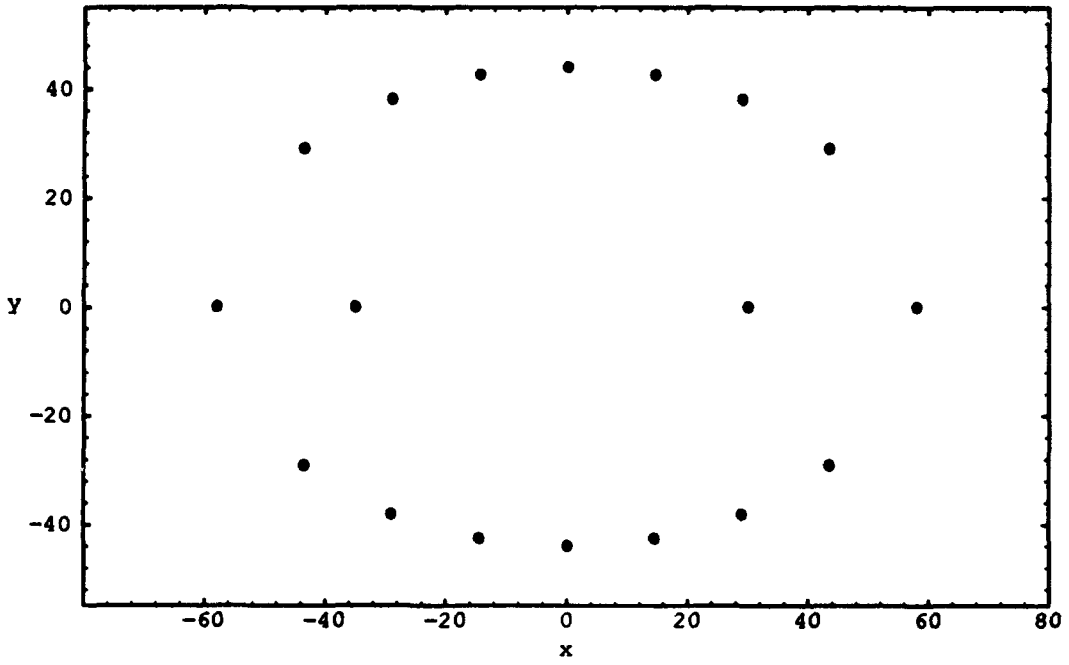


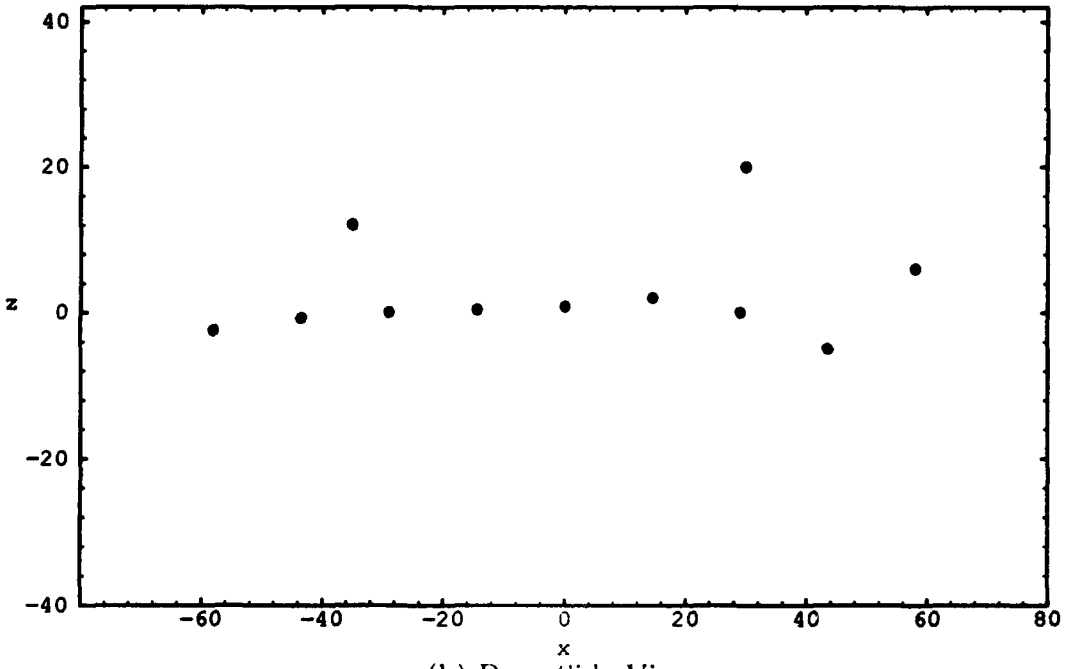
Figure B.18 Spar Chin Position Effects Trend: Trend appearing when eigenvalues of the energy matrix from data capturing the movement in position of the chin are compared. The vertical axis represents the magnitude of the eigenvalues.

Elliptical Points	x	y	z
1	58.0	0.0000	6.0
2	43.5	29.1033	-5.0
3	29.0	38.1051	0.0
4	14.5	42.6028	2.0
5	0.0	44.0000	0.8
6	-14.5	42.6028	0.4
7	-29.0	38.1051	0.0
8	-43.5	29.1033	-0.
9	-58.0	0.0000	-2.5
10	-43.5	-29.1033	-0.8
11	-29.0	-38.1051	0.0
12	-14.5	-42.6028	0.4
13	0.0	-44.0000	0.8
14	14.5	-42.6028	2.0
15	29.0	-38.1051	0.0
16	43.5	-29.1033	-5.0
Nose	30.0	0.0000	20.0
Chin	-35.0	0.0000	12.0

Table B.4 Degg Coordinate Data: This modified egg has 'depth' – the points along the ellipse have been given a height similar to the residual 'height' of points along the seal line after the trend has been removed.



(a) Degg Top View

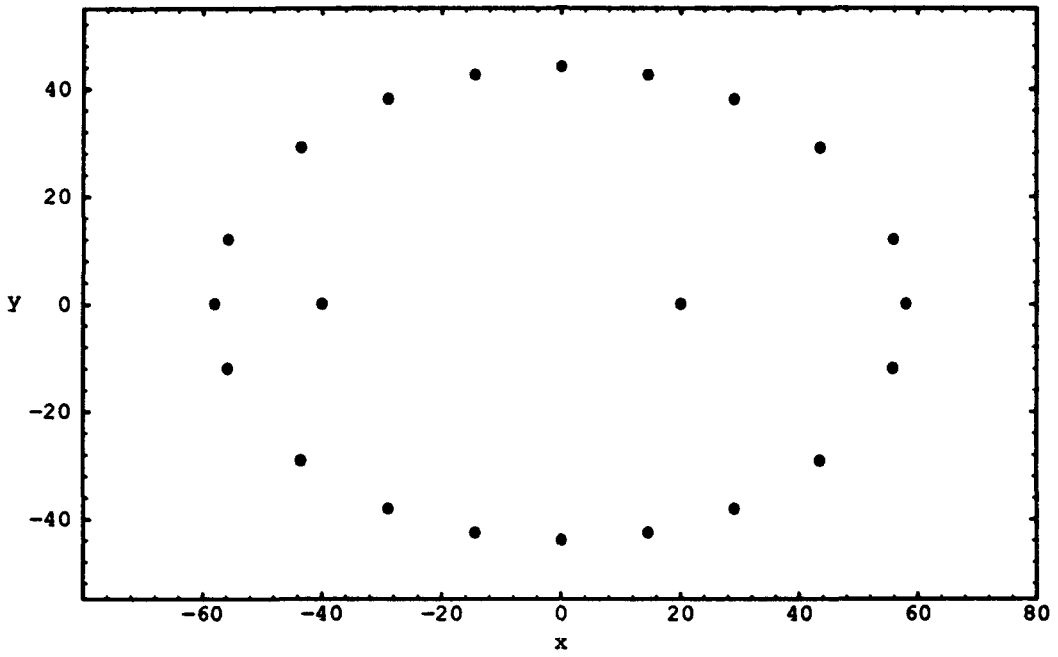


(b) Degg Side View

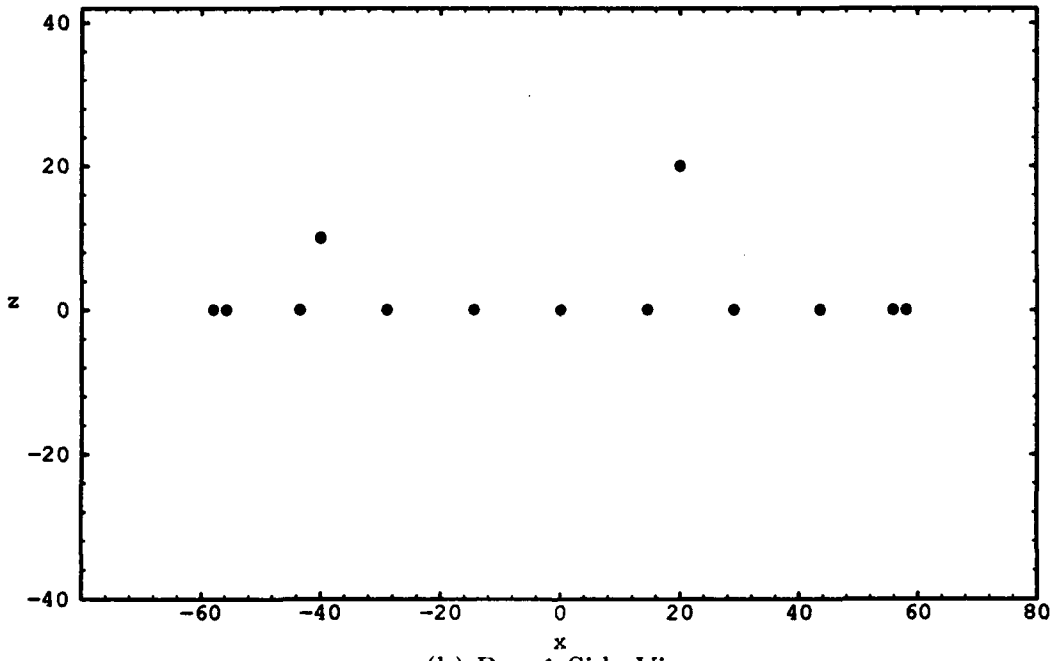
Figure B.19 Degg Two-Dimensional Plots: The Degg which consists of 18 points; (a) Degg top view, (b) Degg side view.

Elliptical Points	x	y	z
1	58.0	0.0000	0.0
2	55.8	12.0000	0.0
3	43.5	29.1033	0.0
4	29.0	38.1051	0.0
5	14.5	42.6028	0.0
6	0.0	44.0000	0.0
7	-14.5	42.6028	0.0
8	-29.0	38.1051	0.0
9	-43.5	29.1033	0.0
10	-55.8	12.0000	0.0
11	-58.0	0.0000	0.0
12	-55.8	-12.0000	0.0
13	-43.5	-29.1033	0.0
14	-29.0	-38.1051	0.0
15	-14.5	-42.6028	0.0
16	0.0	-44.0000	0.0
17	14.5	-42.6028	0.0
18	29.0	-38.1051	0.0
19	43.5	-29.1033	0.0
20	55.8	-12.0000	0.0
Nose	20.0	0.0000	20.0
Chin	-40.0	0.0000	10.0

Table B.5 Begg1 Coordinate Data.



(a) Beggl Top View

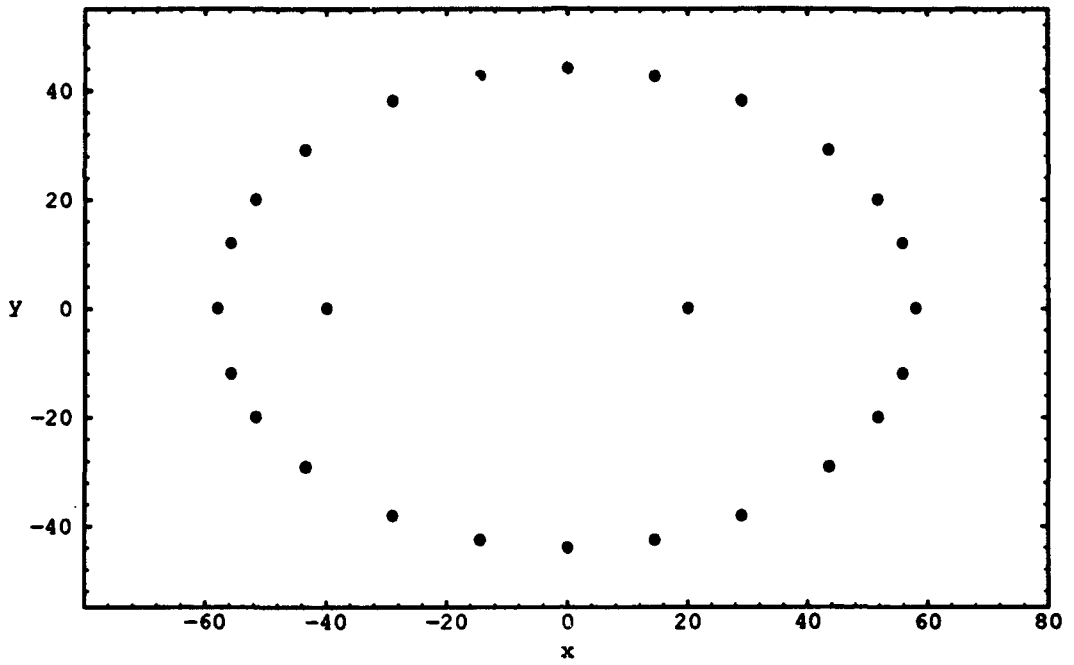


(b) Beggl Side View

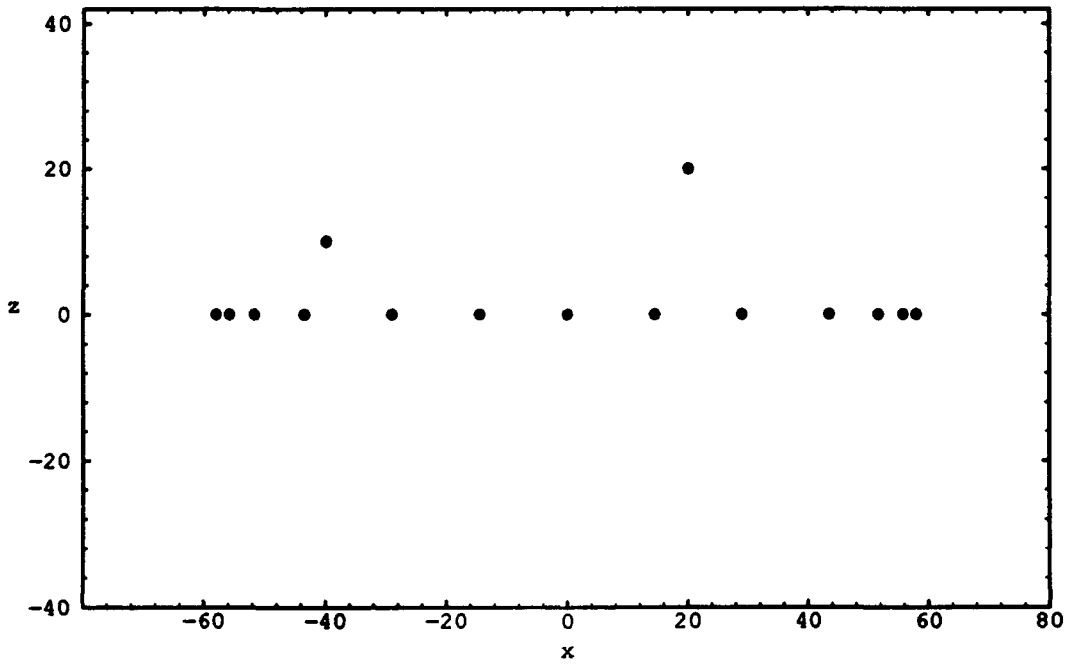
Figure B.20 Beggl Two-Dimensional Plots: The Beggl which consists of 22 points; (a) Beggl top view, (b) Beggl side view.

Elliptical Points	x	y	z
1	58.0	0.0000	0.0
2	55.8	12.0000	0.0
3	51.7	20.0000	0.0
4	43.5	29.1033	0.0
5	29.0	38.1051	0.0
6	14.5	42.6028	0.0
7	0.0	44.0000	0.0
8	-14.5	42.6028	0.0
9	-29.0	38.1051	0.0
10	-43.5	29.1033	0.0
11	-51.7	20.0000	0.0
12	-55.8	12.0000	0.0
13	-58.0	0.0000	0.0
14	-55.8	-12.0000	0.0
15	-51.7	-20.0000	0.0
16	-43.5	-29.1033	0.0
17	-29.0	-38.1051	0.0
18	-14.5	-42.6028	0.0
19	0.0	-44.0000	0.0
20	14.5	-42.6028	0.0
21	29.0	-38.1051	0.0
22	43.5	-29.1033	0.0
23	51.7	-20.0000	0.0
24	55.8	-12.0000	0.0
Nose	20.0	0.0000	20.0
Chin	-40.0	0.0000	10.0

Table B.6 Begg2 Coordinate Data.



(a) Begg2 Top View

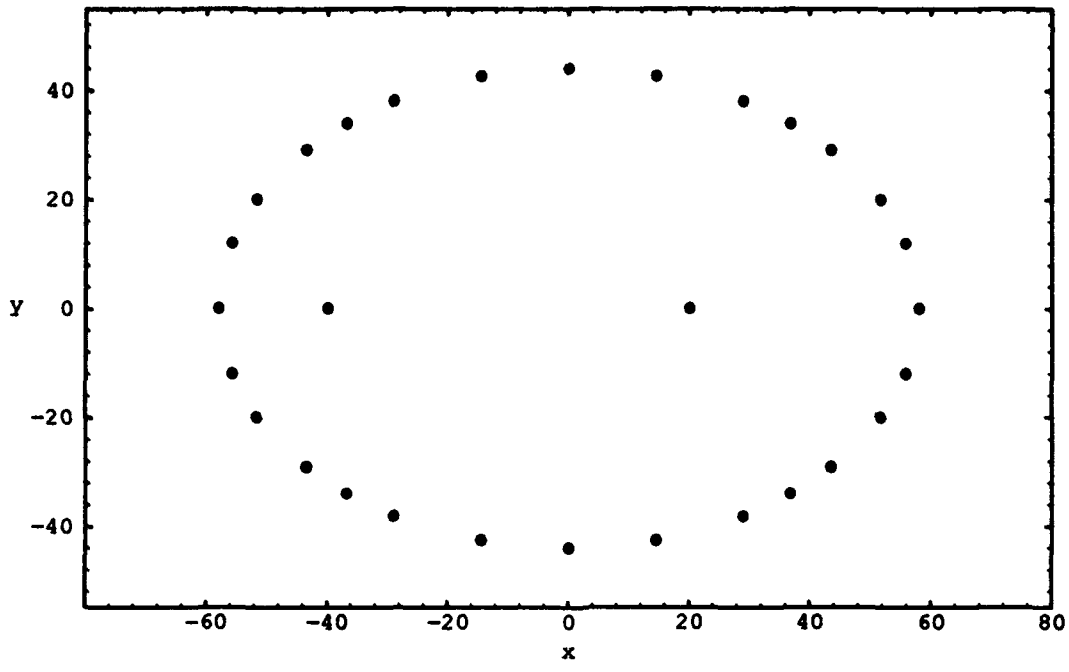


(b) Begg2 Side View

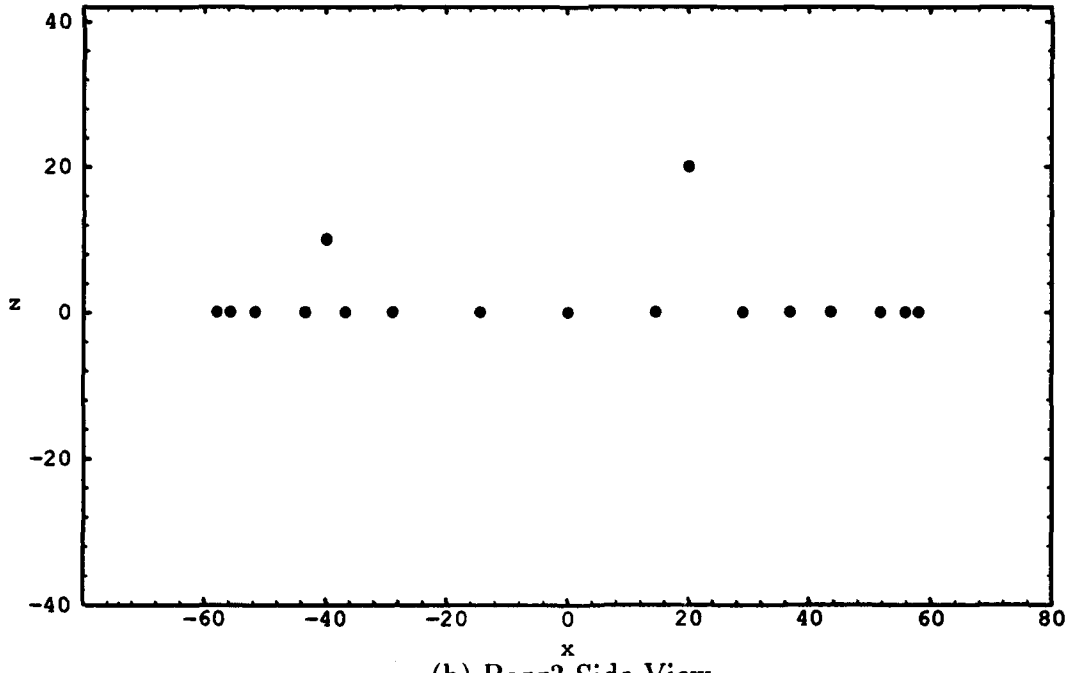
Figure B.21 Begg2 Two-Dimensional Plots: The Begg2 which consists of 26 points; (a) Begg2 top view, (b) Begg2 side view.

Elliptical Points	x	y	z
1	58.0	0.0000	0.0
2	55.8	12.0000	0.0
3	51.7	20.0000	0.0
4	43.5	29.1033	0.0
5	36.8	34.0000	0.0
6	29.0	38.1051	0.0
7	14.5	42.6028	0.0
8	0.0	44.0000	0.0
9	-14.5	42.6028	0.0
10	-29.0	38.1051	0.0
11	-36.8	34.0000	0.0
12	-43.5	29.1033	0.0
13	-51.7	20.0000	0.0
14	-55.8	12.0000	0.0
15	-58.0	0.0000	0.0
16	-55.8	-12.0000	0.0
17	-51.7	-20.0000	0.0
18	-43.5	-29.1033	0.0
19	-36.8	-34.0000	0.0
20	-29.0	-38.1051	0.0
21	-14.5	-42.6028	0.0
22	0.0	-44.0000	0.0
23	14.5	-42.6028	0.0
24	29.0	-38.1051	0.0
25	36.8	-34.0000	0.0
26	43.5	-29.1033	0.0
27	51.7	-20.0000	0.0
28	55.8	-12.0000	0.0
Nose	20.0	0.0000	20.0
Chin	-10.0	0.0000	10.0

Table B.7 Begg3 Coordinate Data.



(a) Begg3 Top View

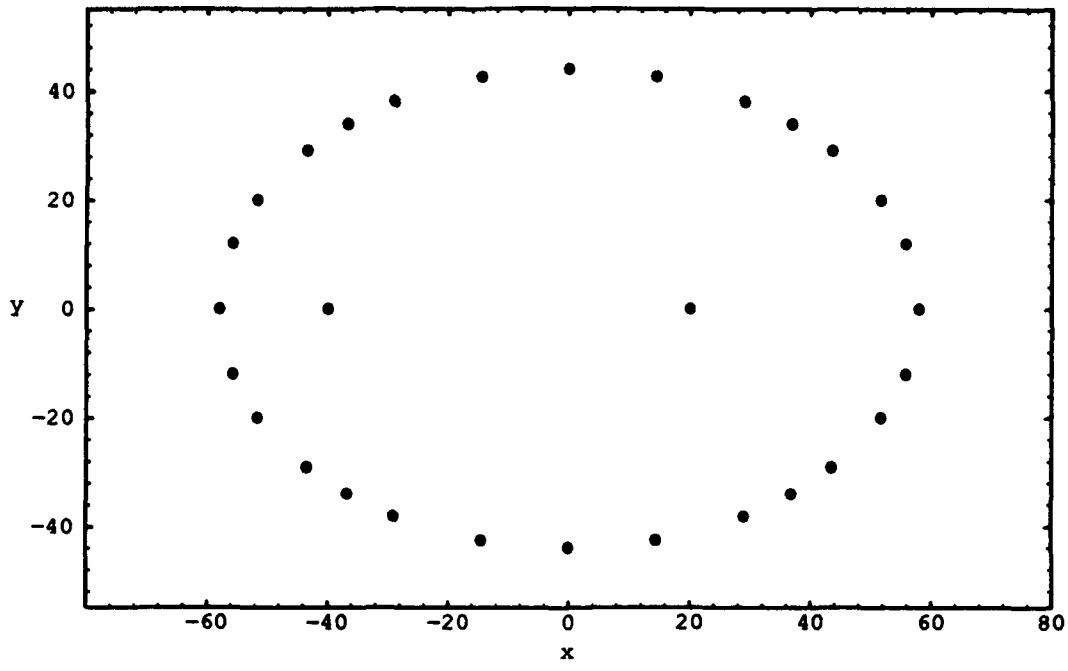


(b) Begg3 Side View

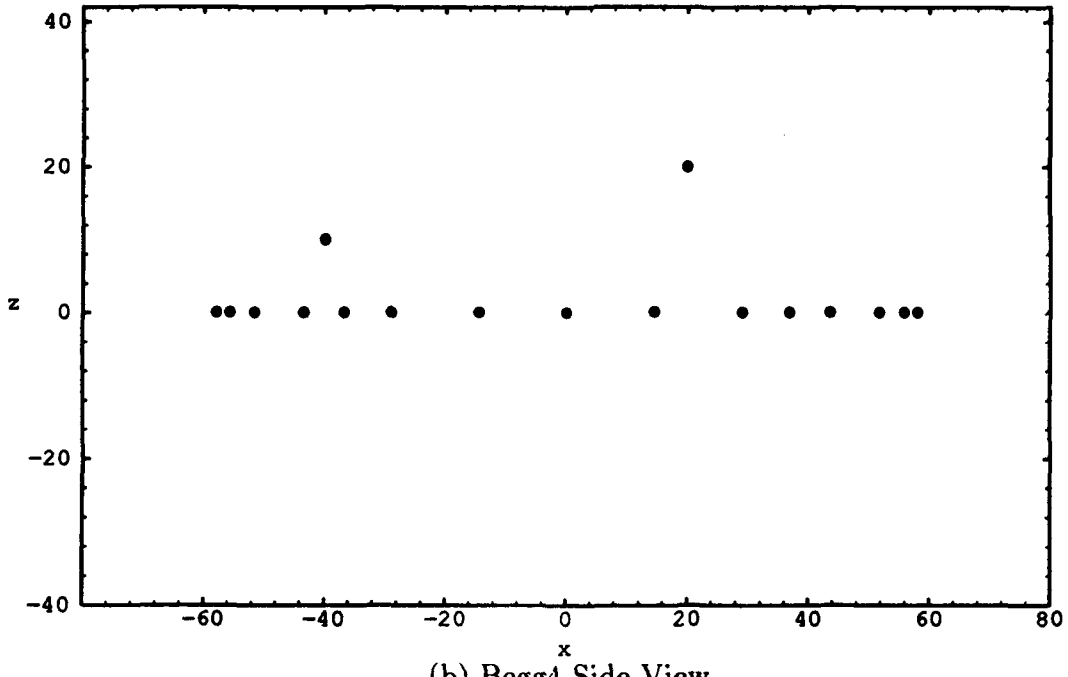
Figure B.22 Begg3 Two-Dimensional Plots: The Begg3 which consists of 30 points; (a) Begg2 top view, (b) Begg2 side view. NOTE: Begg3 and Begg differ in Nose and Chin.

Elliptical Points	x	y	z
1	55.8	12.0000	0.0
2	43.5	29.1033	0.0
3	29.0	38.1051	0.0
4	14.5	42.6028	0.0
5	0.0	44.0000	0.0
6	-14.5	42.6028	0.0
7	-29.0	38.1051	0.0
8	-43.5	29.1033	0.0
9	-55.8	12.0000	0.0
10	-55.8	-12.0000	0.0
11	-43.5	-29.1033	0.0
12	-29.0	-38.1051	0.0
13	-14.5	-42.6028	0.0
14	0.0	-44.0000	0.0
15	14.5	-42.6028	0.0
16	29.0	-38.1051	0.0
17	43.5	-29.1033	0.0
18	55.8	-12.0000	0.0
Nose	20.0	0.0000	20.0
Chin	-40.0	0.0000	10.0

Table B.8 Begg4 Coordinate Data.



(a) Begg4 Top View



(b) Begg4 Side View

Figure B.23 Begg4 Two-Dimensional Plots: The Begg4 which consists of 20 points; (a) Begg4 top view, (b) Begg4 side view.

Appendix C. Data Preparation Steps for Face 0002v

The tables in this appendix show the sequence of steps taken to prepare the data for form analysis. Face 0002v is used as an example.

Table C.1 is the raw data file, as taken by the digitizer (see Section 5.1.1.3). Table C.2 is the data after it has been smoothed by the Pro/ENGINEER CAD program (see Section 5.1.2.1). Table C.3 is the data after it has undergone principal components analysis, which places the origin of the coordinate system at the centroid of the seal area (see Section 5.1.2.2). Table C.4 shows the result of the sparsing operation performed by Pro/ENGINEER (see Section 5.1.2.3). Table C.5 contains the results of the residual calculation for the entire, 'un-sparsed' file, with the residual value, e, for each point given in the third column (see Section 5.2.1).

Note that in all of the tables, except Table C.4.4, the last 14 points are the spacial delineators. In the sparsed data file, the final point shows the overall length of the seal area as the x coordinate and the overall width of the seal area as the y coordinate. The 14 points prior to this are the spacial delineators.

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
1	148.94	88.08	118.34	36	201.47	25.36	81.72
2	150.91	88.45	117.84	37	201.09	20.64	81.31
3	153.34	87.70	117.02	38	201.02	18.45	81.09
4	155.70	86.81	115.50	39	199.84	14.75	82.00
5	157.33	85.86	113.33	40	197.88	7.42	82.04
6	158.58	83.78	111.38	41	197.37	5.45	82.19
7	160.64	80.91	108.29	42	196.53	1.81	82.35
8	161.88	79.07	107.47	43	196.11	-0.19	82.50
9	163.10	77.80	106.39	44	195.90	-2.91	82.29
10	164.86	76.03	105.30	45	195.20	-5.86	82.72
11	166.66	74.75	104.33	46	193.84	-11.43	82.21
12	169.54	73.30	102.65	47	192.87	-13.28	81.88
13	171.38	72.23	101.90	48	190.37	-16.08	83.22
14	173.67	70.96	100.96	49	188.48	-18.14	83.14
15	176.98	70.03	99.41	50	185.33	-20.49	82.72
16	179.87	69.34	98.06	51	183.67	-21.46	83.26
17	182.10	68.64	97.31	52	180.57	-22.79	83.54
18	184.12	67.86	96.73	53	178.09	-23.32	83.75
19	187.39	66.81	95.81	54	176.11	-24.11	84.81
20	189.51	66.14	95.10	55	173.70	-25.28	86.43
21	191.92	65.67	94.23	56	171.35	-26.13	87.55
22	194.01	65.03	93.49	57	168.97	-26.89	88.04
23	195.77	64.25	92.83	58	165.70	-28.03	89.12
24	199.17	61.27	91.09	59	163.71	-28.28	89.27
25	200.90	59.03	89.79	60	159.09	-29.11	89.10
26	202.22	57.32	88.57	61	155.75	-29.31	88.80
27	203.44	55.28	87.25	62	151.88	-29.56	89.49
28	204.20	52.42	86.24	63	148.51	-29.64	89.62
29	204.92	49.93	85.05	64	144.29	-29.18	90.07
30	204.50	46.45	84.44	65	141.25	-28.98	89.74
31	204.02	44.03	84.32	66	138.09	-28.69	89.51
32	203.56	38.38	82.87	67	133.55	-27.68	90.14
33	203.39	35.98	82.32	68	129.96	-26.40	90.45
34	202.51	32.79	82.46	69	127.99	-25.42	90.37
35	202.22	30.24	81.93	70	125.34	-23.88	90.35

Table C.1 Raw Digitized Data Points.

(continued on following page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
71	123.16	-22.19	89.79	106	110.77	60.26	99.57
72	120.46	-20.34	88.88	107	112.78	61.30	100.26
73	117.65	-17.61	88.43	108	115.78	62.28	101.25
74	116.65	-15.55	88.80	109	118.38	62.98	102.15
75	115.18	-14.05	87.76	110	121.13	64.46	102.99
76	113.97	-12.34	86.91	111	123.30	65.82	103.59
77	112.60	-8.61	86.64	112	125.39	67.03	104.37
78	110.70	-6.05	85.26	113	127.69	68.12	105.37
79	109.97	-3.17	84.83	114	129.06	69.56	105.89
80	108.43	-0.70	84.02	115	131.18	71.31	106.79
81	107.14	2.91	83.72	116	132.85	73.02	107.50
82	105.65	5.84	83.97	117	134.55	75.10	108.18
83	104.17	8.65	84.22	118	135.88	75.66	109.89
84	103.69	11.17	84.73	119	137.08	77.98	110.75
85	102.40	14.17	84.75	120	138.13	80.56	112.04
86	101.55	16.51	84.81	121	138.73	82.36	112.95
87	101.27	18.90	85.08	122	139.04	84.33	113.25
88	100.32	21.67	85.12	123	140.98	85.29	115.63
89	98.70	24.95	84.98	124	142.93	86.58	117.00
90	97.95	27.64	85.48	125	145.25	87.43	117.98
91	96.60	31.39	85.54	126	147.35	88.24	118.18
92	95.86	33.27	85.63	127	149.08	88.43	118.16
93	95.01	37.79	86.39	128	136.87	73.66	112.22
94	95.44	39.76	87.22	129	148.92	78.75	122.65
95	95.43	41.72	87.87	130	161.65	75.67	111.38
96	95.39	44.22	89.10	131	128.80	60.76	107.81
97	95.72	46.07	90.30	132	151.58	56.24	131.22
98	96.59	47.85	91.59	133	173.10	64.77	105.03
99	97.07	50.30	92.59	134	128.92	46.00	104.18
100	98.48	51.66	93.90	135	152.63	34.09	113.04
101	100.01	53.11	94.96	136	173.34	48.32	102.37
102	101.78	55.17	96.01	137	123.54	14.48	103.36
103	103.70	56.72	96.79	138	185.44	19.66	100.58
104	106.12	57.81	97.90	139	156.48	-0.86	110.16
105	107.88	58.62	98.51	140	158.86	-15.87	114.95
				141	159.93	-26.69	105.74

Table C.1.b. Raw Digitized Data Points.

(continued from previous page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
1	149.08	88.43	118.16	36	201.59	25.42	81.65
2	150.91	88.45	117.84	37	201.19	21.48	81.37
3	153.32	87.65	116.79	38	200.65	17.95	81.47
4	155.46	86.79	115.28	39	199.58	13.54	81.71
5	157.20	85.48	113.40	40	198.36	9.21	82.08
6	158.85	83.51	111.00	41	197.26	4.90	82.19
7	160.37	81.25	109.05	42	196.67	2.36	82.35
8	161.88	79.26	107.38	43	196.18	-0.43	82.38
9	163.28	77.63	106.39	44	195.74	-2.99	82.51
10	164.88	76.19	105.34	45	194.98	-6.74	82.41
11	167.02	74.69	104.09	46	193.97	-10.19	82.27
12	169.20	73.43	102.96	47	192.36	-13.60	82.44
13	171.53	72.16	101.84	48	190.58	-15.83	82.75
14	174.01	71.07	100.76	49	188.06	-18.24	83.03
15	176.84	70.11	99.48	50	185.83	-20.03	83.04
16	179.65	69.34	98.26	51	183.19	-21.58	83.18
17	182.03	68.61	97.37	52	180.78	-22.53	83.52
18	184.54	67.77	96.62	53	178.26	-23.41	84.03
19	187.01	66.94	95.88	54	175.97	-24.24	85.00
20	189.61	66.21	95.05	55	173.72	-25.17	86.27
21	191.81	65.62	94.27	56	171.34	-26.10	87.34
22	193.90	64.99	93.52	57	168.67	-27.02	88.24
23	196.32	63.52	92.47	58	166.13	-27.73	88.81
24	198.61	61.52	91.24	59	162.84	-28.47	89.17
25	200.77	59.21	89.82	60	159.52	-28.90	89.06
26	202.19	57.21	88.54	61	155.57	-29.33	89.13
27	203.29	55.01	87.36	62	152.04	-29.51	89.30
28	204.19	52.54	86.18	63	148.23	-29.46	89.73
29	204.54	49.60	85.25	64	144.69	-29.27	89.81
30	204.48	46.80	84.60	65	141.21	-28.95	89.77
31	204.03	42.95	83.88	66	137.63	-28.45	89.80
32	203.66	39.47	83.17	67	133.87	-27.59	90.03
33	203.16	35.72	82.55	68	130.50	-26.50	90.32
34	202.71	33.01	82.24	69	127.77	-25.23	90.39
35	202.07	29.46	82.04	70	125.50	-23.83	90.17

Table C.2 Smoothed Data Points.

(continued on following page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
71	122.99	-22.14	89.67	106	110.48	60.06	99.45
72	120.43	-20.05	89.04	107	113.11	61.28	100.36
73	118.25	-17.83	88.71	108	115.65	62.19	101.22
74	116.49	-15.74	88.33	109	118.43	63.24	102.13
75	115.26	-13.98	87.82	110	120.94	64.42	102.91
76	113.91	-11.66	87.10	111	123.27	65.77	103.65
77	112.42	-9.00	86.27	112	125.46	66.99	104.44
78	111.09	-5.94	85.58	113	127.38	68.24	105.21
79	109.70	-3.31	84.70	114	129.31	69.66	106.02
80	108.52	-0.32	84.19	115	131.03	71.29	106.73
81	107.07	2.68	83.91	116	132.86	73.14	107.49
82	105.65	5.80	83.97	117	134.43	74.59	108.52
83	104.50	8.55	84.31	118	135.84	76.25	109.61
84	103.42	11.33	84.57	119	137.03	78.07	110.90
85	102.55	13.95	84.76	120	137.98	80.30	111.91
86	101.74	16.53	84.88	121	138.63	82.42	112.75
87	101.05	19.03	85.00	122	139.58	83.99	113.94
88	100.10	21.84	85.06	123	140.98	85.40	115.29
89	98.99	24.75	85.19	124	143.05	86.43	116.87
90	97.75	28.00	85.33	125	145.18	87.42	117.72
91	96.80	30.77	85.55	126	147.35	88.24	118.18
92	95.82	34.15	85.86	127	149.08	88.43	118.16
93	95.44	36.94	86.42	128	136.87	73.66	112.22
94	95.29	39.76	87.16	129	148.92	78.75	122.65
95	95.42	41.90	88.07	130	161.65	75.67	111.38
96	95.51	44.00	89.09	131	128.80	60.76	107.81
97	95.90	46.05	90.33	132	151.58	56.24	131.22
98	96.46	48.07	91.50	133	173.10	64.77	105.03
99	97.38	49.94	92.69	134	128.92	46.00	104.18
100	98.52	51.69	93.82	135	152.63	34.09	113.04
101	100.09	53.31	94.96	136	173.34	48.32	102.37
102	101.83	55.00	95.92	137	123.54	14.48	103.36
103	103.87	56.57	96.90	138	185.44	19.66	100.58
104	105.90	57.72	97.73	139	156.48	-0.86	110.16
105	108.26	58.90	98.66	140	158.86	-15.87	114.95
				141	159.93	-26.69	105.74

Table C.2.b. Smoothed Data Points.

(continued from previous page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
1	58.68	4.20	-10.92	36	-15.03	51.14	9.70
2	58.46	6.03	-10.68	37	-18.88	50.37	9.19
3	57.25	8.37	-9.91	38	-22.26	49.49	8.39
4	55.90	10.44	-8.70	39	-26.40	47.99	7.29
5	54.08	12.09	-7.20	40	-30.43	46.35	6.10
6	51.52	13.60	-5.32	41	-34.51	44.84	5.15
7	48.77	14.93	-3.94	42	-36.90	44.00	4.50
8	46.35	16.27	-2.78	43	-39.56	43.24	3.91
9	44.43	17.54	-2.20	44	-41.99	42.55	3.29
10	42.66	19.01	-1.54	45	-45.59	41.43	2.64
11	40.75	21.03	-0.71	46	-48.89	40.10	2.11
12	39.08	23.09	0.04	47	-52.03	38.17	1.32
13	37.41	25.32	0.79	48	-53.98	36.16	0.63
14	35.90	27.70	1.52	49	-56.04	33.42	-0.03
15	34.44	30.45	2.45	50	-57.58	31.03	-0.32
16	33.18	33.20	3.37	51	-58.83	28.25	-0.66
17	32.08	35.52	4.00	52	-59.46	25.75	-1.09
18	30.87	37.95	4.45	53	-59.98	23.15	-1.67
19	29.69	40.34	4.90	54	-60.38	20.77	-2.69
20	28.57	42.87	5.46	55	-60.83	18.41	-4.03
21	27.63	45.03	6.01	56	-61.30	15.93	-5.17
22	26.67	47.06	6.53	57	-61.76	13.17	-6.13
23	24.81	49.34	7.15	58	-62.11	10.56	-6.73
24	22.39	51.46	7.85	59	-62.46	7.20	-7.09
25	19.66	53.41	8.68	60	-62.60	3.86	-6.94
26	17.32	54.66	9.46	61	-62.64	-0.11	-6.93
27	14.83	55.57	10.12	62	-62.46	-3.64	-6.99
28	12.10	56.25	10.73	63	-61.99	-7.45	-7.24
29	9.01	56.33	11.02	64	-61.46	-10.95	-7.14
30	6.16	56.02	11.08	65	-60.85	-14.38	-6.89
31	2.30	55.21	11.02	66	-60.03	-17.89	-6.66
32	-1.21	54.53	11.01	67	-58.80	-21.56	-6.56
33	-4.95	53.68	10.87	68	-57.38	-24.81	-6.48
34	-7.62	52.98	10.64	69	-55.89	-27.41	-6.18
35	-11.05	52.00	10.14	70	-54.36	-29.53	-5.58

Table C.3 Oriented Data Points.

(continued on following page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
71	-52.59	-31.85	-4.65	106	30.63	-36.55	3.16
72	-50.45	-34.18	-3.49	107	31.78	-33.84	2.41
73	-48.17	-36.12	-2.62	108	32.61	-31.24	1.65
74	-46.05	-37.66	-1.75	109	33.57	-28.39	0.86
75	-44.33	-38.71	-0.85	110	34.66	-25.80	0.24
76	-42.10	-39.81	0.39	111	35.91	-23.36	-0.30
77	-39.55	-41.02	1.81	112	37.07	-21.09	-0.92
78	-36.59	-42.03	3.17	113	38.27	-19.07	-1.50
79	-34.08	-43.14	4.62	114	39.65	-17.03	-2.07
80	-31.17	-44.02	5.79	115	41.23	-15.18	-2.50
81	-28.18	-45.16	6.74	116	43.03	-13.19	-2.94
82	-25.00	-46.27	7.37	117	44.52	-11.52	-3.72
83	-22.14	-47.16	7.66	118	46.23	-9.98	-4.50
84	-19.29	-47.97	8.02	119	48.16	-8.64	-5.44
85	-16.61	-48.59	8.40	120	50.46	-7.50	-6.01
86	-14.01	-49.14	8.85	121	52.64	-6.67	-6.42
87	-11.49	-49.60	9.27	122	54.33	-5.60	-7.31
88	-8.65	-50.27	9.83	123	55.86	-4.09	-8.40
89	-5.68	-51.09	10.34	124	57.01	-1.97	-9.81
90	-2.38	-52.02	10.92	125	57.95	0.22	-10.53
91	0.45	-52.70	11.31	126	58.65	2.46	-10.91
92	3.89	-53.35	11.75	127	58.68	4.20	-10.92
93	6.76	-53.48	11.79	128	44.15	-9.25	-7.63
94	9.67	-53.37	11.64	129	50.19	3.00	-17.28
95	11.93	-53.05	11.19	130	43.70	15.61	-7.42
96	14.19	-52.78	10.61	131	31.40	-18.44	-5.63
97	16.40	-52.22	9.80	132	29.81	3.29	-30.38
98	18.57	-51.49	9.06	133	30.73	26.09	-3.92
99	20.55	-50.42	8.23	134	16.26	-19.67	-5.11
100	22.39	-49.14	7.45	135	4.36	2.58	-17.19
101	24.06	-47.45	6.60	136	14.13	24.79	-4.69
102	25.75	-45.57	5.93	137	-14.12	-28.05	-10.55
103	27.30	-43.42	5.21	138	-15.24	34.10	-9.32
104	28.41	-41.30	4.55	139	-30.62	3.09	-21.70
105	29.54	-38.86	3.79	140	-44.46	3.90	-29.56
				141	-57.00	4.12	-22.81

Table C.3.b. Oriented Data Points.

(continued from previous page)

Pt. No.	x coord	y coord	z coord	Pt. No.	x coord	y coord	z coord
1	58.68	4.20	-10.92	28	3.89	-53.35	11.75
2	58.46	6.03	-10.68	29	11.93	-53.05	11.19
3	55.90	10.44	-8.70	30	18.57	-51.49	9.06
4	51.52	13.60	-5.32	31	24.06	-47.45	6.60
5	44.43	17.54	-2.20	32	28.41	-41.30	4.55
6	37.41	25.32	0.79	33	37.07	-21.09	-0.92
7	30.87	37.95	4.45	34	41.23	-15.18	-2.50
8	26.67	47.06	6.53	35	46.23	-9.98	-4.50
9	22.39	51.46	7.85	36	50.46	-7.50	-6.01
10	14.83	55.57	10.12	37	54.33	-5.60	-7.31
11	9.01	56.33	11.02	38	57.01	-1.97	-9.81
12	-7.62	52.98	10.64	39	58.65	2.46	-10.91
13	-22.26	49.49	8.39	40	44.15	-9.25	-7.63
14	-41.99	42.55	3.29	41	50.19	3.00	-17.28
15	-48.89	40.10	2.11	42	43.70	15.61	-7.42
16	-53.98	36.16	0.63	43	31.40	-18.44	-5.63
17	-58.83	28.25	-0.66	44	29.81	3.29	-30.38
18	-60.38	20.77	-2.69	45	30.73	26.09	-3.92
19	-61.76	13.17	-6.13	46	16.26	-19.67	-5.11
20	-62.46	7.20	-7.09	47	4.36	2.58	-17.19
21	-62.46	-3.64	-6.99	48	14.13	24.79	-4.69
22	-60.03	-17.89	-6.66	49	-14.12	-28.05	-10.55
23	-55.89	-27.41	-6.18	50	-15.24	34.10	-9.32
24	-50.45	-34.18	-3.49	51	-30.62	3.09	-21.70
25	-44.33	-38.71	-0.85	52	-44.46	3.90	-29.56
26	-31.17	-44.02	5.79	53	-57.01	4.12	-22.81
27	-22.14	-47.16	7.66	54	121	110	0

Table C.4 Sparsed Data Points.

Pt. No.	x coord	y coord	e coord	Pt. No.	x coord	y coord	e coord
1	58.68	4.20	-5.49	36	-15.03	51.14	0.99
2	58.46	6.03	-5.29	37	-18.88	50.37	1.00
3	57.25	8.37	-4.44	38	-22.26	49.49	0.79
4	55.90	10.44	-3.20	39	-26.40	47.99	0.69
5	54.08	12.09	-1.57	40	-30.43	46.35	0.53
6	51.52	13.60	0.56	41	-34.51	44.84	0.42
7	48.77	14.93	2.20	42	-36.90	44.00	0.19
8	46.35	16.27	3.50	43	-39.56	43.24	-0.10
9	44.43	17.54	4.10	44	-41.99	42.55	-0.48
10	42.66	19.01	4.65	45	-45.59	41.43	-0.74
11	40.75	21.03	5.16	46	-48.89	40.10	-0.76
12	39.08	23.09	5.46	47	-52.[illegible]	38.17	-0.67
13	37.41	25.32	5.60	48	-53.[illegible]8	36.16	-0.34
14	35.90	27.70	5.53	49	-56.04	33.42	0.37
15	34.44	30.45	5.37	50	-57.58	31.03	1.20
16	33.18	33.20	5.03	51	-58.83	28.25	2.13
17	32.08	35.52	4.49	52	-59.46	25.75	2.81
18	30.87	37.95	3.62	53	-59.98	23.15	3.28
19	29.69	40.34	2.66	54	-60.38	20.77	3.14
20	28.57	42.87	1.59	55	-60.83	18.41	2.57
21	27.63	45.03	0.67	56	-61.30	15.93	2.13
22	26.67	47.06	-0.27	57	-61.76	13.17	1.84
23	24.81	49.34	-1.30	58	-62.11	10.56	1.77
24	22.39	51.46	-2.17	59	-62.46	7.20	1.93
25	19.66	53.41	-2.84	60	-62.60	3.86	2.45
26	17.32	54.66	-3.02	61	-62.64	-0.11	2.65
27	14.83	55.57	-3.04	62	-62.46	-3.64	2.58
28	12.10	56.25	-2.92	63	-61.99	-7.45	2.12
29	9.01	56.33	-2.56	64	-61.46	-10.95	1.82
30	6.16	56.02	-2.10	65	-60.85	-14.38	1.48
31	2.30	55.21	-1.29	66	-60.03	-17.89	0.92
32	-1.21	54.53	-0.60	67	-58.80	-21.56	0.01
33	-4.95	53.68	0.06	68	-57.38	-24.81	-0.95
34	-7.62	52.98	0.45	69	-55.89	-27.41	-1.58
35	-11.05	52.00	0.77	70	-54.36	-29.53	-1.79

Table C.5 Residual Data Points.

(continued on following page)

Pt. No.	x coord	y coord	e coord	Pt. No.	x coord	y coord	e coord
71	-52.59	-31.85	-1.84	106	30.63	-36.55	1.35
72	-50.45	-34.18	-1.74	107	31.78	-33.84	2.13
73	-48.17	-36.12	-1.79	108	32.61	-31 24	2.76
74	-46.05	-37.66	-1.68	109	33.57	-28.39	3.35
75	-44.33	-38.71	-1.29	110	34.66	-25.80	3.83
76	-42.10	-39.81	-0.61	111	35.91	-23.36	4.18
77	-39.55	-41.02	0.16	112	37.07	-21.09	4.30
78	-36.59	-42.03	0.99	113	38.27	-19.07	4.27
79	-34.08	-43.14	1.77	114	39.65	-17.03	4.15
80	-31.17	-44.02	2.40	115	41.23	-15.18	4.00
81	-28.18	-45.16	2.59	116	43.03	-13.19	3.78
82	-25.00	-46.27	2.42	117	44.52	-11.52	3.13
83	-22.14	-47.16	2.03	118	46.23	-9.98	2.35
84	-19.29	-47.97	1.73	119	48.16	-8.64	1.29
85	-16.61	-48.59	1.58	120	50.46	-7.50	0.44
86	-14.01	-49.14	1.52	121	52.64	-6.67	-0.29
87	-11.49	-49.60	1.50	122	54.33	-5.60	-1.40
88	-8.65	-50.27	1.38	123	55.86	-4.09	-2.64
89	-5.68	-51.09	1.04	124	57.01	-1.97	-4.12
90	-2.38	-52.02	0.62	125	57.95	0.22	-4.94
91	0.45	-52.70	0.21	126	58.65	2.46	-5.44
92	3.89	-53.35	-0.21	127	58.68	4.20	-5.45
93	6.76	-53.48	-0.53	128	44.15	-9.25	-0.24
94	9.67	-53.37	-0.84	129	50.19	3.00	-9.99
95	11.93	-53.05	-1.23	130	43.70	15.61	-0.51
96	14.19	-52.78	-1.80	131	31.40	-18.44	1.48
97	16.40	-52.22	-2.35	132	29.81	3.29	-19.65
98	18.57	-51.49	-2.70	133	30.73	26.09	1.50
99	20.55	-50.42	-2.82	134	16.26	-19.67	3.50
100	22.39	-49.14	-2.73	135	4.36	2.58	-4.03
101	24.06	-47.45	-2.37	136	14.13	24.79	2.94
102	25.75	-45.57	-1.74	137	-14.12	-28.05	-3.78
103	27.30	-43.42	-0.98	138	-15.24	34.10	-5.59
104	28.41	-41.30	-0.22	139	-30.62	3.09	-8.67
105	29.54	-38.86	0.59	140	-44.46	3.90	-17.73
				141	-57.01	4.12	-12.58

Table C.5.b. Residual Data Points.

(continued from previous page)

Appendix D. Survey Questions and Answers

The following are the questions contained in the survey discussed in Section 5.5 and the answers provided by the subjects.

1. Because you ordered a custom MBU-5/P mask, the standard issue sizes must not have fit. Why? (i.e. air leaks around bridge of nose and eye area, comfort reasons....)

- Cluster 1: The other masks were too large for my face causing leaks around the nose and eyes; also I had to tighten the mask to the point of discomfort.
- Cluster 3: Air leaked around side of nose approximately $\frac{1}{4}$ inch below bridge of nose, unless tightened so tight that bridge of nose VERY painful and cause tears to eyes.
- Cluster 5: Air leaks around bridge of nose and eye area. Had to be tightened so much in that area it was painful and it still had to be readjusted/refitted every flight or every other flight to prevent leaks.
- Cluster 6: My first mask was the old 2 piece green mask due to the shape of my nose. I had continual leaks beneath both eyes. To stop these made the mask too tight for comfort. I hoped a custom would solve that, but I was sadly disappointed.

2. Do you use the custom mask made for you - or did you decide to wear a standard mask?

- Cluster 1: I had been using the custom mask even though a lot of times when I went into the chamber the mask would still have a leak.
- Cluster 3: I use my custom mask.
- Cluster 5: Custom Mask.

- Cluster 6: The custom is too large and bulky. It tends to obstruct your field of view and is incompatible with eyeglasses. (The mask rides too high on the bridge of the nose. NOTE: This is true of all MBU-5/P masks.)

3. Does the custom mask you currently wear fit better than the standard mask(s) you have been fitted with? (i.e. standard mask leaked around the nose-eye area, new mask does not leak...)

- Cluster 1: The custom fit around the nose eye area but at times it was painful to wear.
- Cluster 3: Yes, see question 1.
- Cluster 5: Yes. Custom mask does not leak and is much more comfortable.
- Cluster 6: I currently wear the new one piece grey mask. It fits better than any I've worn. It still won't seal under pressure at 41,000 feet, but I'll never see that altitude in a B-1B. (This subject is referring to the MBU-12/P mask)

4. What is it about the custom mask you currently wear that you do not like? (eg., the latex is uncomfortable in warm weather)

- Cluster 1: Hurts to wear. Isn't always reliable.
- Cluster 3: I like the mask under all conditions, in warm weather I use a little powder.
- Cluster 5: Nothing. It is leaps and bound above the standard mask. Its slipperiness during hot weather is easily prevented by using talcum powder.
- Cluster 6: Don't currently wear it. Hate the whole thing.

5. How long have you been using the custom mask you current wear?

- Cluster 1: 1.5 years.
- Cluster 3: I have been using custom masks since pilot training - my first assignment.

- Cluster 5: Over 2.5 years.
- Cluster 6: Used it for three months of flying before I gave it up.

The following questions pertain to the test mask sent to the survey subjects.

6. How did you test the fit of this mask?

   - Cluster 1: I took a chamber flight to FL 350 with positive pressure breathing starting at FL 280.
   - Cluster 3: Breathed normally and under pressure with oxygen mask tester.
   - Cluster 5: The mask was connected to MQ1 oxygen mask tester and was tested for leaks.
   - Cluster 6: The mask was connected to the tester.

7. Was a good seal obtained?

   - Cluster 1: Yes.
   - Cluster 3: No.
   - Cluster 5: No.
   - Cluster 6: No. It was like holding a balloon to my face. Once pressure was applied, it inflated making it hard to see over.

8. Were there any leaks? If so, where?

   - Cluster 1: No.
   - Cluster 3: Yes, mask leaked below the bridge of nose.
   - Cluster 5: Yes. Around bridge of nose and eye area.
   - Cluster 6: Yes. Around the bridge of the nose.

9. Was the mask comfortable?

   - Cluster 1: YES!

- Cluster 3: No.
- Cluster 5: No. It leaked into my eyes and was very big.
- Cluster 6: There were no pressure points. However, the sheer bulk of this mask would make it unbearable to fly with.

10. Would you be able to wear this mask to fly? (your opinion)

- Cluster 1: Yes.
- Cluster 3: Not at all.
- Cluster 5: No. Large size of mask reduced visibility. Material around bridge of nose cut into field of vision. And again, could not obtain a good seal.
- Cluster 6: No.

11. If a standard mask was available in this size, would you wear it?

- Cluster 1: Yes.
- Cluster 3: Sure, less problems procuring new one when needed.
- Cluster 5: No.
- Cluster 6: No.

12. Any other comments?

- Cluster 1: This mask seemed more round on the outside unlike my regular mask. So when I went to get a mask seal I didn't have to tighten it down all the way on my face. It was extremely comfortable. It also gives me hope that I can go through chamber flights without discomfort. Thanks.
- Cluster 3: NOTE: In a conversation with cluster 3 survey subject, the subject mentioned he has always required a custom mask due to the shape of his nose. His nose had been broken several times before entering pilot training.

- Cluster 5: It was noted by both life support technicians and myself that the mask was lopsided at the bridge of the nose. One side was not straight.
- Cluster 6: The test mask also began to lose its shape as it was fitted to my face. The rubber would bend then crease causing leaks. It also rode way too high up on the bridge of my nose. I wouldn't even be able to wear glasses with the test mask.

Appendix E. Cluster Representatives

This appendix shows a comparison between the faces closest to the centroid of a cluster (prototype seals) and the faces tested with masks based on the prototype seals, referred to as survey seals. Figures E.1 through E.4 are the graphical representations of this comparison. Each face is shown in front and left side views. The front views are shown with the nose facing out of the page, the bridge of the nose to the right and the chin toward the left. The left side views are shown as if the wearer is laying face up. In the side views, the point with the largest z value is the tip of the nose and the point with the second highest z value is the chin. The (a) and (b) plots are the survey seals and the (c) and (d) plots are the prototype seals. The balance of this appendix are photographs of the actual plaster face casts from which the prototype seals are taken.

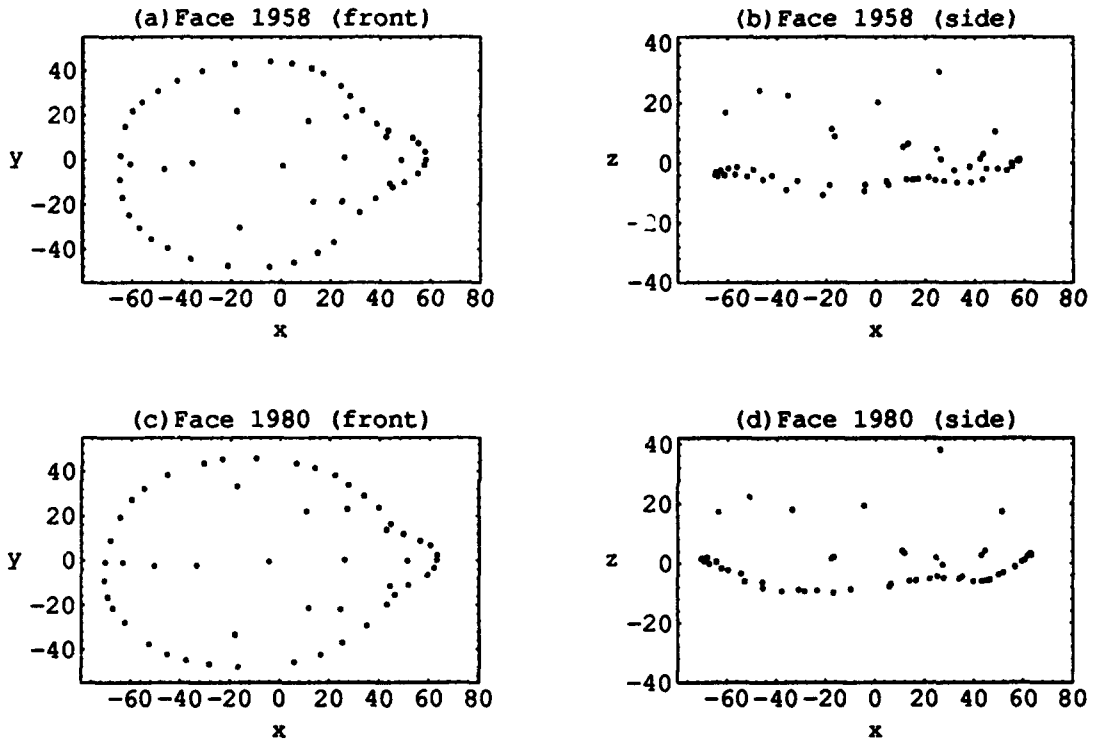


Figure E.1 Comparison between prototype seal 1980 and survey seal 1958

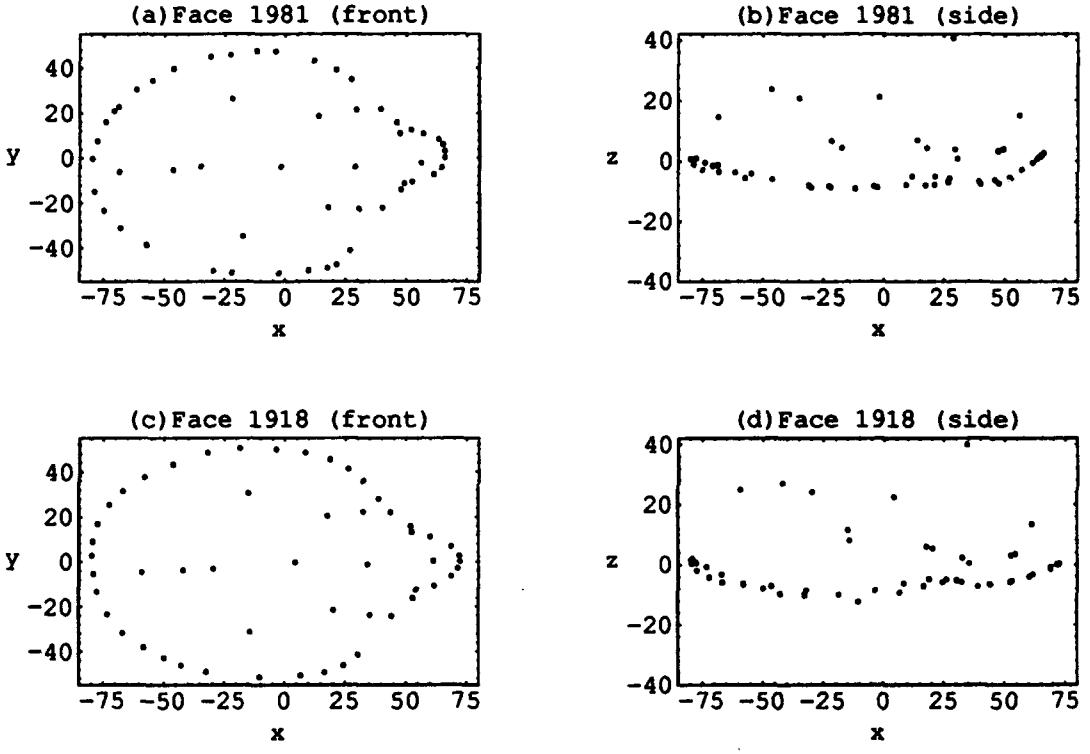


Figure E.2 Comparison between prototype seal 1918 and survey seal 1981

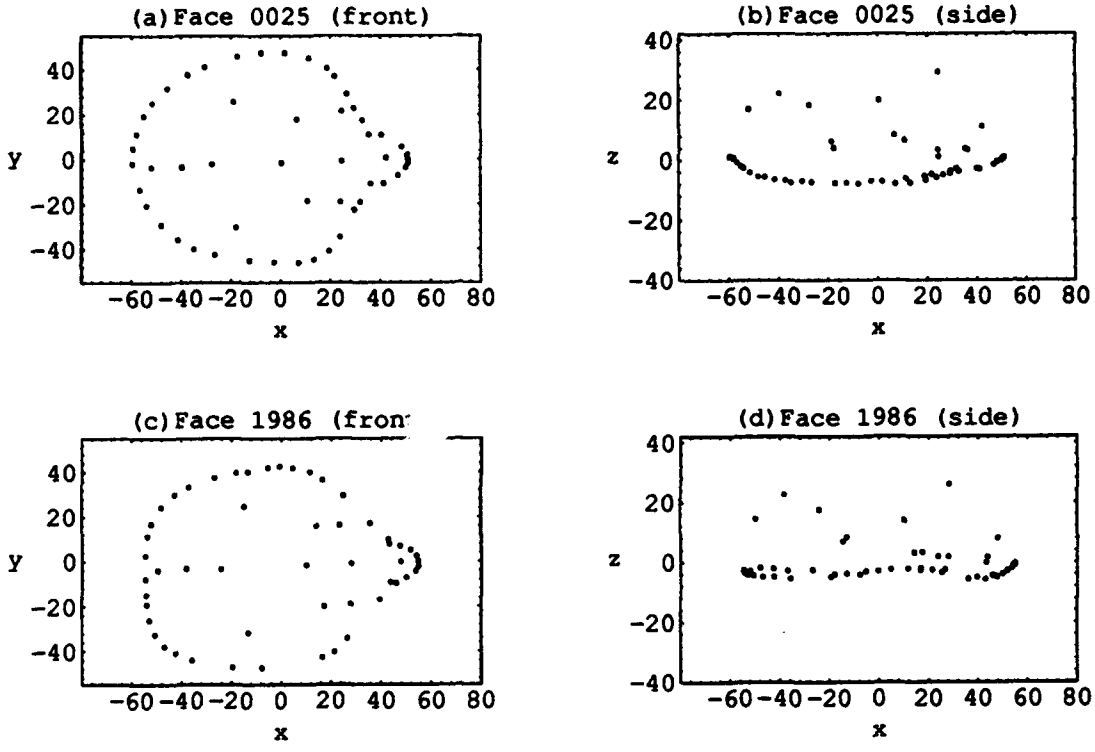


Figure E.3 Comparison between prototype seal 1986 and survey seal 0025

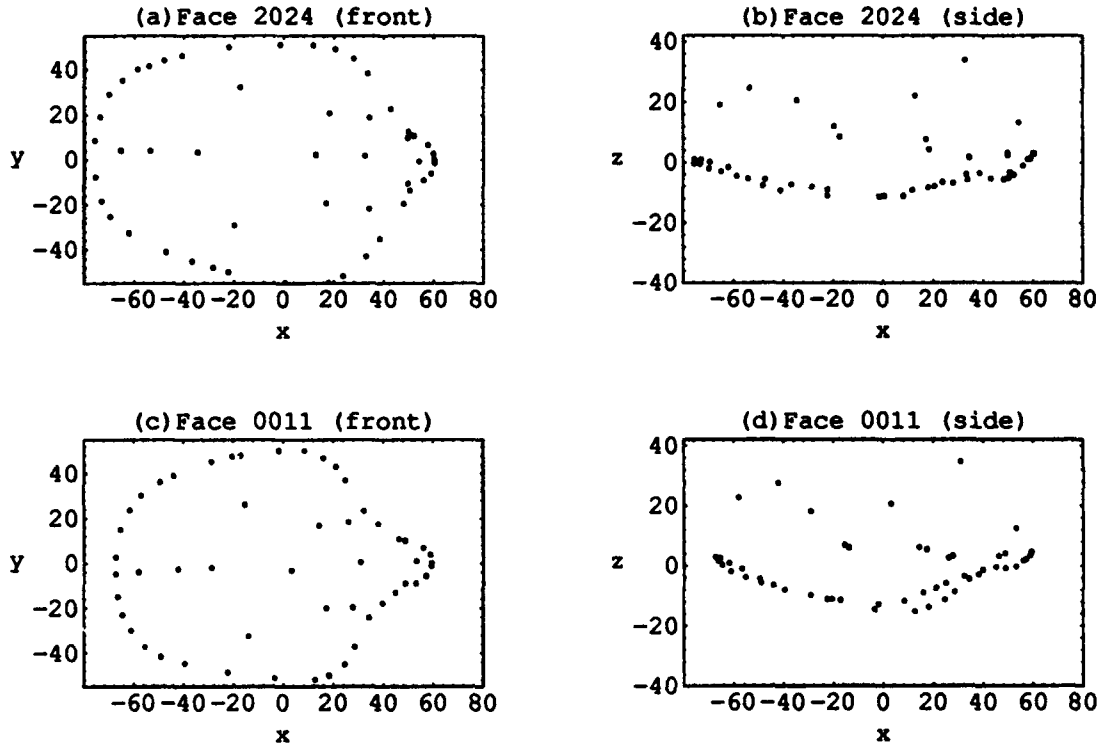


Figure E.4 Comparison between prototype seal 0011 and survey seal 2024

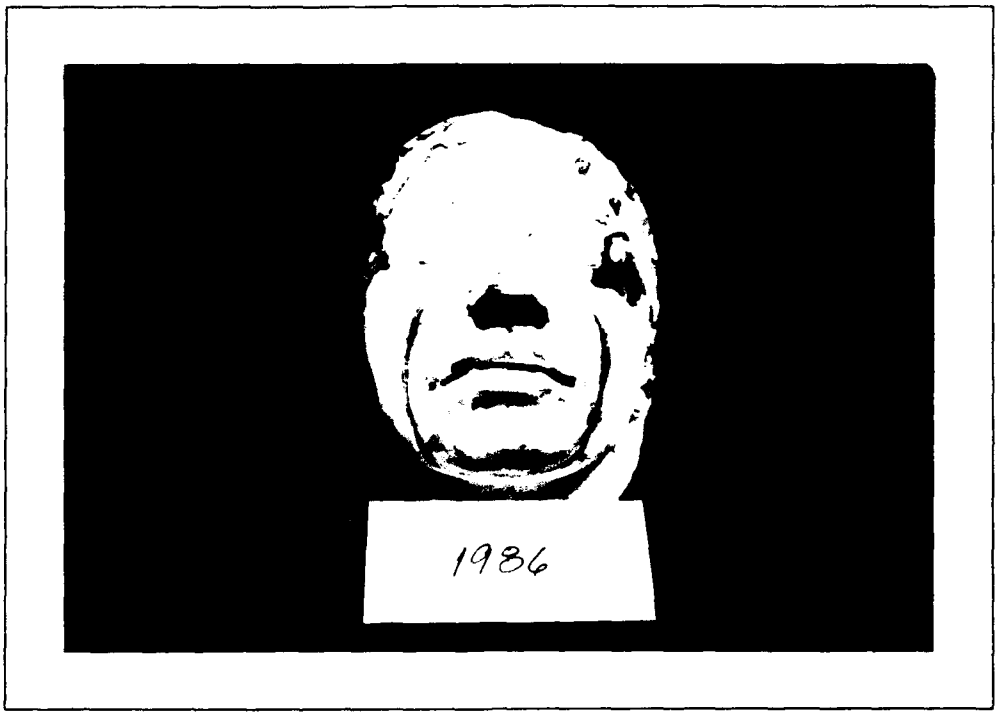


Figure E.5 Prototype face 1986 (front)

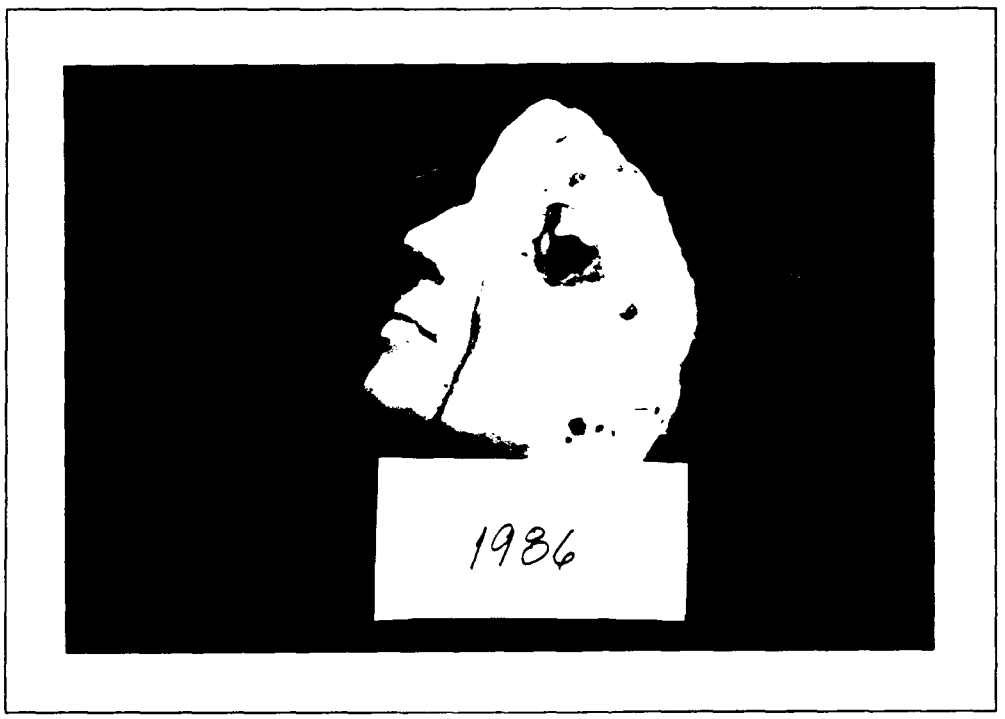


Figure E.6 Prototype face 1986 (side)

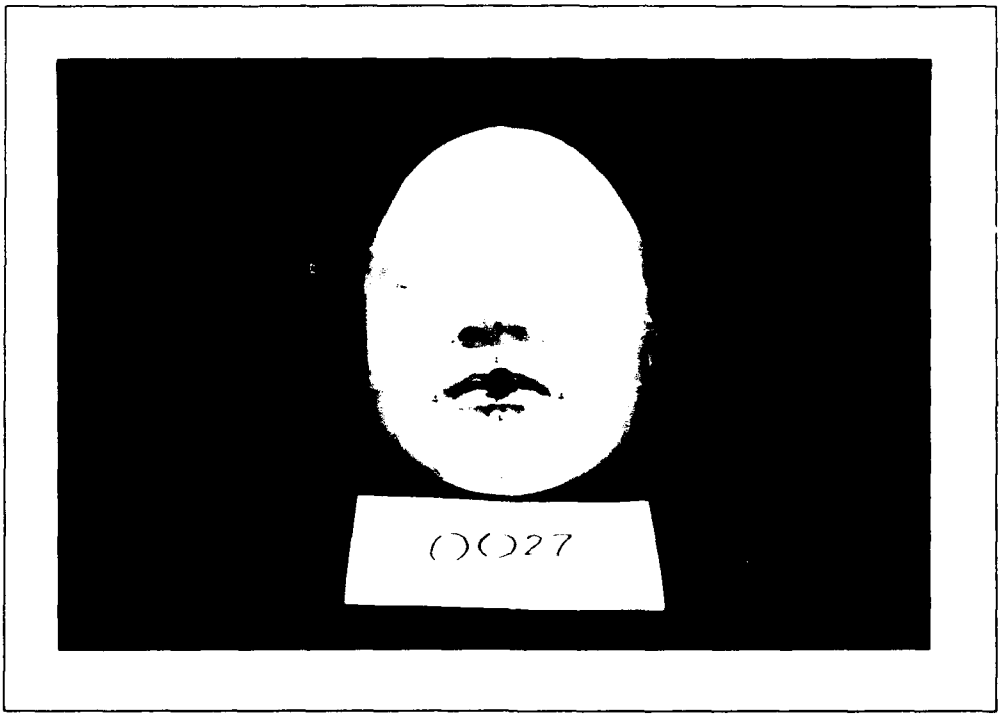


Figure E.7 Prototype face 0027 (front)

Figure E.8 Prototype face 0027 (side)

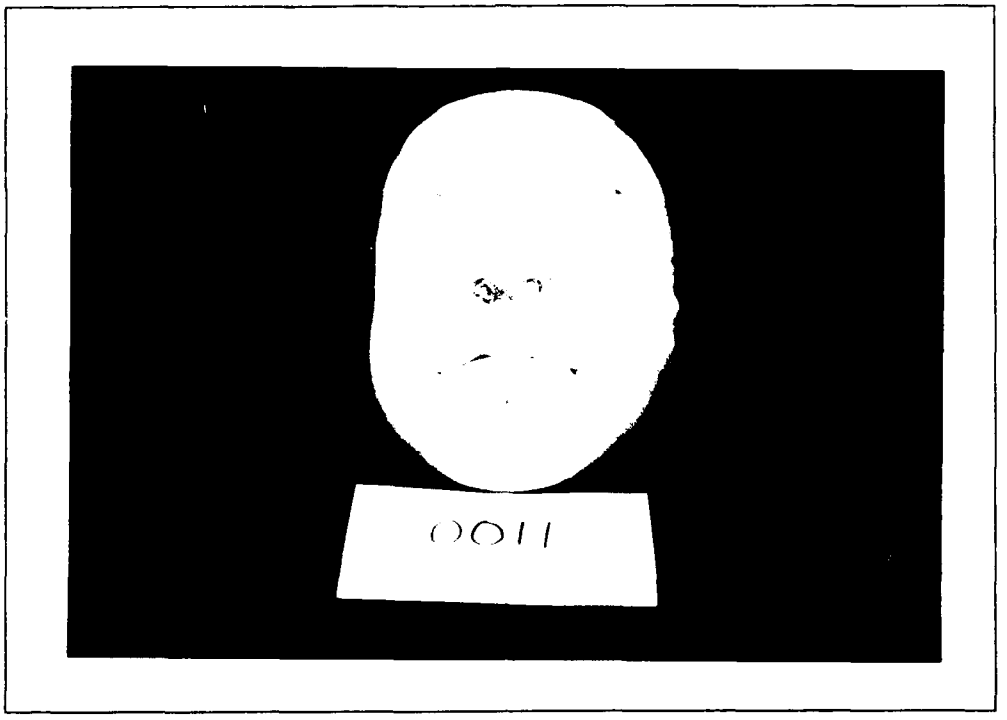


Figure E.9 Prototype face 0011 (front)

Figure E.10 Prototype face 0011 (side)

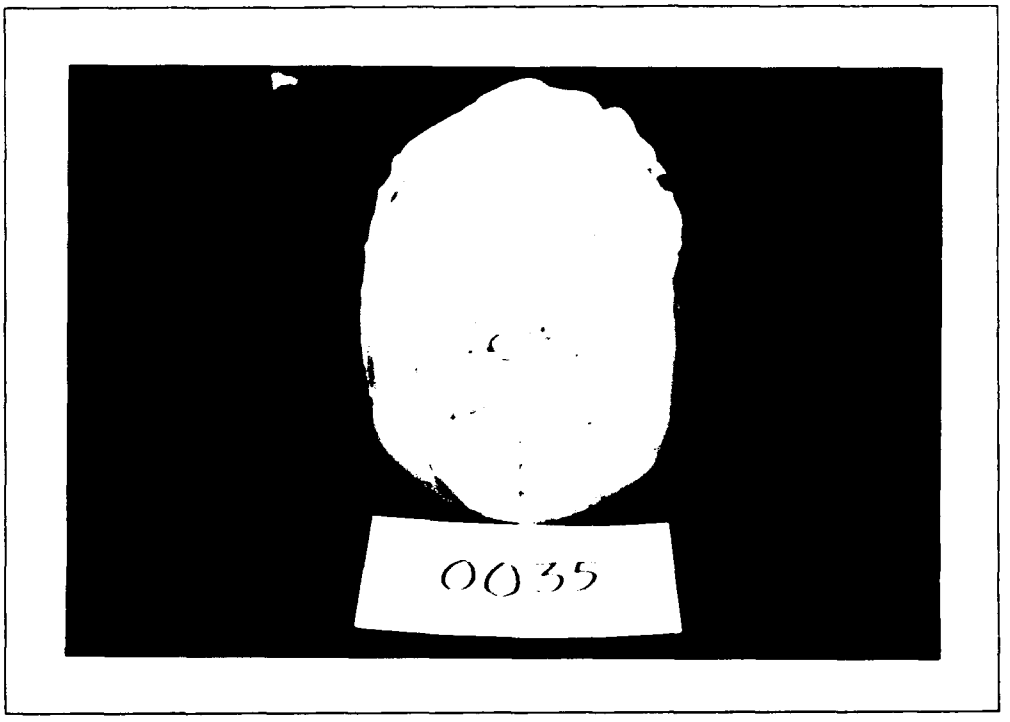


Figure E.11 Prototype face 0035 (front)

Figure E.12 Prototype face 0035 (side)

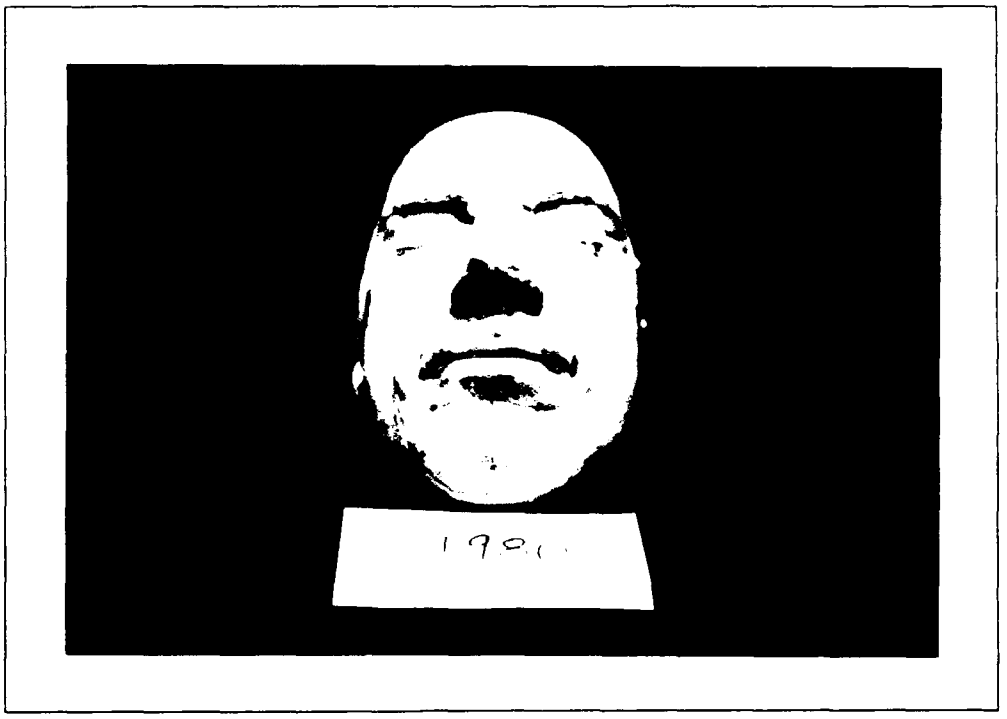


Figure E.13 Prototype face 1980 (front)

Figure E.14 Prototype face 1980 (side)

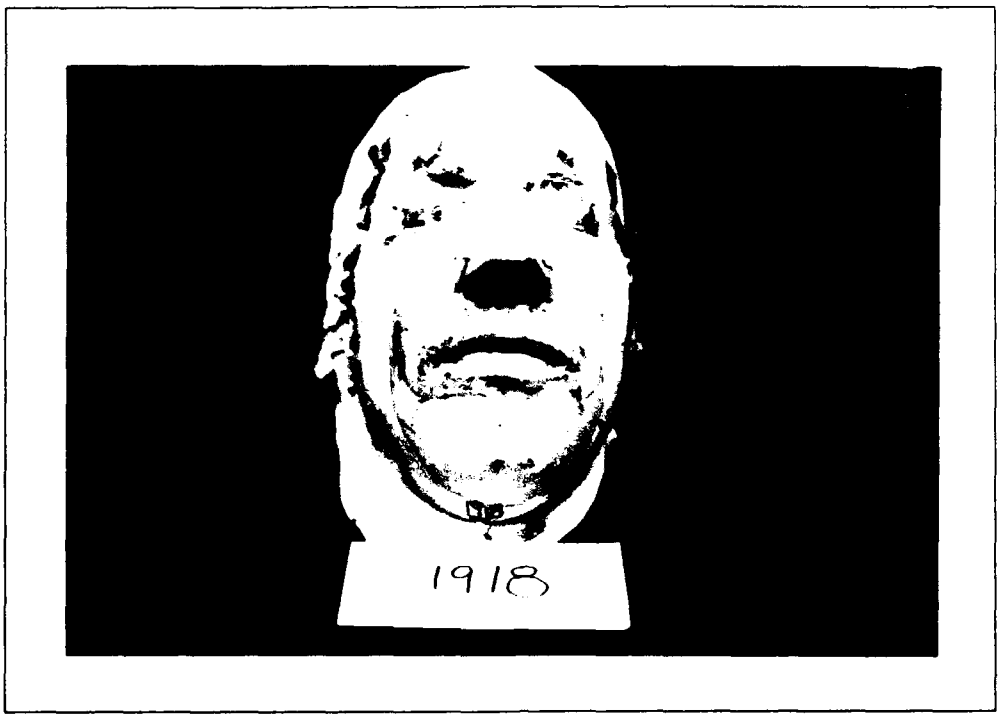


Figure E.15 Prototype face 1918 (front)

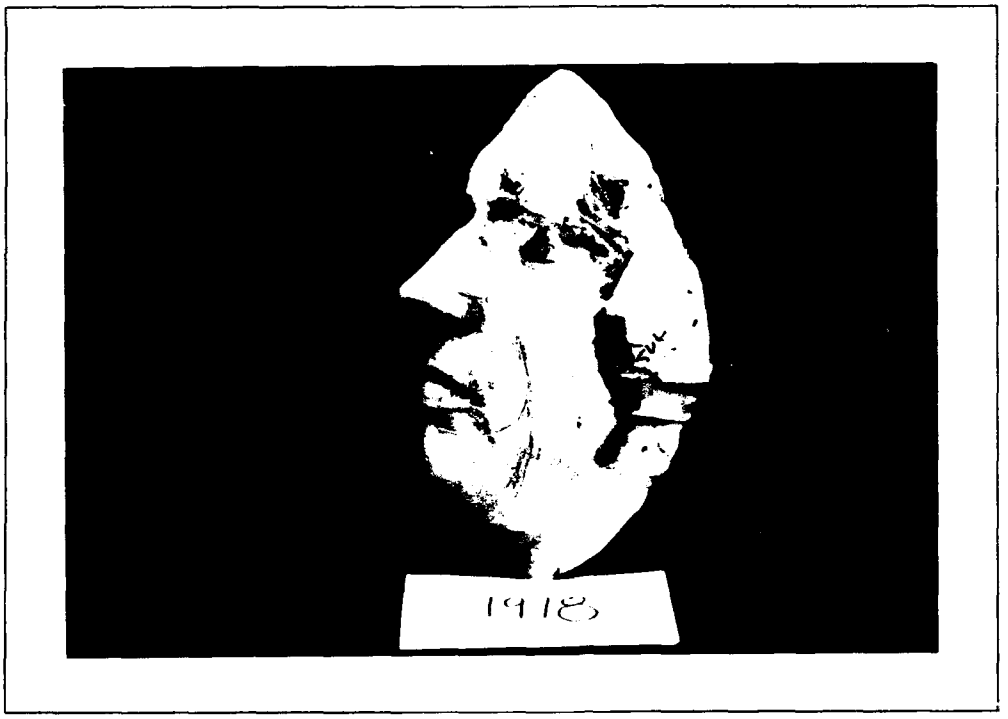


Figure E.16 Prototype face 1918 (side)

Bibliography

1. Bezdek, James C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
2. Bezdek, James C. and Sankar K. Pal. *Fuzzy Models for Pattern Recognition: Methods That Search for Structures in Data*. New York: Institute of Electronics and Electrical Engineers, 1992.
3. Bookstein, Fred. L. "Foundations of Morphometrics," *Annual Review Ecol. Syst.*, *13*:451–470 (1982).
4. Bookstein, Fred L. "Principal Warps: Thin Plate Splines and the Decomposition of Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(6):567–589 (June 1989).
5. Bookstein, Fred L. *Morphometric Tools for Landmark Data*. New York: Cambridge University Press, 1991.
6. Brodkin, Christopher. *The Application of Kriging for Controlled Minimization of Large Data Sets*. MS thesis, Air Force Institute of Technology, December 1991.
7. Bulfinch, Thomas. *Bulfinch's Mythology*. New York: Thomas Y. Crowell Company, 1947.
8. Churchill, Edmund and Gilbert S. Daniels. *Nomographs of Head Measurements*. Technical Report WADC-TR 53-14, WADC, May 1953.
9. Churchill, Edmund and Bruce Truett. *Metrical Relations Among Dimensions of the Head and Face*. Technical Report WADC-TR 56-621, WADC, June 1957.
10. Clark, Isobel. *Practical Geostatistics*. Ripple Road, Barking, Essex England: Applied Science Publishers LTD, 1979.
11. Clauser, Charles, et al. *Measurer's Handbook: U.S. Army Anthropometric Survey*. TR 88/043, Natick, Massachusetts 01760-5000: U.S. Army Natick Research, Development and Engineering Center, 1988.
12. (CMS), Custom Mask Shop. Informal discussions and interviews with various members of the CMS., January – October 1993.
13. Cooke, Julian P. *Human Compatibility Testing of a Pressure-Breathing Mask, MBU-12/P*. Technical Report SAM-TR 76-11, SAM, May 1976.
14. Corner, Brian D. and Joan T. Richtsmeier. "Morphometric Analysis of Craniofacial Growth in Cebus Apella," *American Journal of Physical Anthropology*, *84*:323–342 (1991).
15. Cover, T. M. "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications to Pattern Recognition," *IEEE Transactions on Electronic Computers*, *EC-14*(3) (1965).

16. Cressie, Noel. "Kriging Nonstationary Data," *Journal of the American Statistical Association*, *81*(395):625–634 (September 1986).

17. Cressie, Noel. "Geostatistics," *The American Statistician*, *43*(4):197–202 (November 1989).

18. Cressie, Noel. *Statistics for Spatial Data*. New York: Wiley-Interscience, 1991.

19. David, Michel. *Developments in Geomathematics 2: Geostatistical Ore Reserve Estimation*. New York: Elsevier Scientific Publishing Company, Inc., 1977.

20. Davis, John C. *Statistics and Data Analysis in Geology* (second Edition). New York: John Wesley & Sons, 1986.

21. Duda, Richard O. and Peter E. Hart. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

22. Emanuel, Irvin, et al. *Anthropometric Sizing and Fit Test of the MC-1 Oral-Nasal Oxygen Mask*. Technical Report WADC-TR 58-505, WADC, March 1959.

23. Faro Technologies, Inc., Industrial Division, 125 Technology Park, Lake Mary, FL 32746-6204. *Industrial Metrecom Manual, Model IND-2*, 1992.

24. Flury, Bernhard and Hans Riedwyl. *Multivariate Statistics: A Practical Approach*. New York: Chapman and Hall, 1988.

25. Gentex Western Operations. *Mask Custom Fit Meeting*. Briefing Slides. 11525 8th Street, Rancho Cucamonga, CA 91730, 1 April 1993.

26. Gonzalez, Rafael C. and Paul Wintz. *Digital Image Processing*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1987.

27. Grant, Michael. *The Application of Kriging in the Statistical Analysis of Anthropometric Data*. MS thesis, Air Force Institute of Technology, March 1990.

28. Hertzberg, H. T. E., et al. *Anthropometry of Flying Personnel*. Technical Report WADC-TR 52-321, WADC, September 1954.

29. Isaaks, Edward H. and R. Mohan Srivastava. *An Introduction to Applied Geostatistics*. New York: Oxford University Press, Inc., 1989.

30. Johnson, Richard A. and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., 1982.

31. Journel, Andre G. *Fundamentals of Geostatistics in Five Lessons*. Washington, D.C.: American Geophysical Union, 1989.

32. Kanel, Laveen and B. Chandrasekaran. "On Dimensionality and Sample Size in Statistical Pattern Recognition," *Pattern Recognition*, *3* (1971).

33. Kerbs, Lynda. "GEO – Statistics: The Variogram," *COGS Computer Contributions*, *12*(2):54–59 (August 1986).

34. Krogman, Wilton Marion. *The Human Skeleton in Forensic Medicine*. 301-327 East Lawrence Avenue, Springfield, Illinois, USA: Charles C. Thomas, 1962. third printing.

35. Lele, Subhash. "Some Comments on Coordinate-Free and Scale-Invariant Methods in Morphometrics," *American Journal of Physical Anthropology*, *85*:407–417 (1991).

36. Lele, Subhash and Joan T. Richtsmeier. "Euclidean Distance Matrix Analysis: A Coordinate-Free Approach for Comparing Biological Shapes Using Landmark Data," *American Journal of Physical Anthropology*, *86*:415–427 (1991).

37. Lele, Subhash and Joan T. Richtsmeier. "On Comparing Biological Shapes: Detection of Influential Landmarks," *American Journal of Physical Anthropology*, *87*:49–65 (1992).

38. McCloskey, Kathy, et al. *Subjective Responses to Positive Breathing Under Sustained High-G Using the COMBAT EDGE System*. Technical Report AAMRL-TR 90-056, AAMRL, August 1990.

39. McConville, J. T. and Milton Alexander. "Anthropometric Sizing Program for Oral-Nasal Oxygen Masks Based on 1967 Air Force Survey Data," *Aviation, Space, and Environmental Medicine* (November 1975).

40. McGee, Donald W. *The Application of Statistical Kriging to Improve Satellite Imagery Resolution*. MS thesis, Air Force Institute of Technology, December 1991.

41. Parametric Technology Corporation, 128 Technology Drive, Waltham, MA 02154. *Pro/ENGINEER, Version 10.0*, 1992.

42. Priedeman, Bill. Telephone interview. Chemical/Material Expert. Stratasys, Inc., Eden Prairie, MN, 4 June 1993.

43. Radoy, Charles H. *Pattern Recognition By Fourier Series Transforms*. MS thesis, Air Force Institute of Technology, March 1967.

44. Raudys, Serunas J. and Amil K. Jain. "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (March 1991).

45. Richtsmeier, Joan T. "Comparative Study of Normal, Crouzon, and Apert Craniofacial Morphology Using Finite Element Scaling Analysis," *American Journal of Physical Anthropology*, *74*:473–493 (1987).

46. Richtsmeier, Joan T., et al. "Advances in Anthropological Morphometrics," *Annual Review Anthropol.*, *21*:283–305 (1992).

47. Ruskai, Mary Beth, et al. *Wavelets and Their Applications*. Boston: Jones and Bartlett Publishers, 1992.

48. Seeler, Henry W. *Development of Oral-Nasal Masks, Oxygen, MC-1 and MBU-5/P*. Technical Report ASD-TR 61-395, ASD, August 1961.

49. Stewart, G. W. *Introduction to Matrix Computations*. London, U.K.: Academic Press, Inc., 1973.

50. Strang, Gilbert. *Linear Algebra and Its Applications*. San Diego: Harcourt Brace Jovanovich, 1988.

51. Stratasys, Inc., 14950 Martin Dr., Eden Prairie, MN 55344. *3D MODELER Operator's Manual*, November 1992.

52. Thieman, David A., "Reverse-Engineering Design Procedures for Creating the MBU-5/12/20P Oxygen Mask." Southern Ohio Council for Higher Education (SOCHE) Contractor Report, June 1993.

53. Toennies, Klaus D., et al. "Registration of 3D Objects and Surfaces," *IEEE Computer Graphics and Applications*, *10*:52–62 (May 1990).

54. Tou, Julius T. and Rafael C. Gonzalez. *Pattern Recognition Principles*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1974.

55. Turk, Matthew A. and Alex P. Pentland. "Face Recognition Using Eigenfaces." *Proceedings of the 1991 IEEE Computer Society Conference On Computer Vision and Pattern Recognition*. 586–591. Boston: IEEE Computer Society Press, 1991.

56. USAF Aerospace Medical Research Laboratory, Anthropology Branch, Wright-Patterson AFB, OH 45433-5300. *USAF Anthropometric Survey*, 1967.

57. USAF Medical Center, Aerospace Physiological Training Directorate. *Operations. CMS Adminstrative Policies*. SOI 55-2. November 1991.

58. USAF Medical Center, Aerospace Physiological Training Directorate. *Operations. CMS Dipping Procedures*. SOI 55-4. November 1991.

59. USAF Medical Center, Aerospace Physiological Training Directorate. *Operations. CMS Mask Mold Curing Process*. SOI 55-5. November 1991.

60. USAF Medical Center, Aerospace Physiological Training Directorate. *Operations. CMS Mold Construction*. SOI 55-3. November 1991.

61. USAF Medical Center, Aerospace Physiological Training Directorate. *Operations. CMS Pressing Hard-Shells*. SOI 55-6. November 1991.

62. Xie, Xuanli Lisa and Gerardo Beni. "A Validity Measure for Fuzzy Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*:841–847 (1991).

63. Zadeh, L. A. "Fuzzy Sets," *Information and Control*, *8* (1965).

64. Zimmerman, H. J. *Fuzzy Set Theory - and Its Applications*. Boston: Kluwer Academic Publishers, 1991.

Vita

Captain Martin E. Piccus was born 8 February 1967 in Cleveland, Ohio. He graduated from Orange High School in Orange Village, Ohio. He attended The Ohio State University, graduating with a Bachelor of Science in Aeronautical and Astronautical Engineering in June 1985. He then entered the U. S. Air Force and was stationed at Los Angeles AFB, California. He worked in the Defense Support Program where he was responsible for the early on-orbit testing of the satellite. He then worked in the Titan Program Office where he was responsible for the performance of the vehicle until he entered the Graduate Systems Engineering Program, School of Engineering, Air Force Institute of Technology in May 1992.

Permament Address: 1 Madison Ct
Beachwood, Ohio 44122

Mr. Gary Smith was born 27 June 1959 in Detroit, Michigan. He graduated from Stanford University in 1981 with a Bachelors of Science in Mechanical Engineering. Mr. Smith worked in the nuclear power, automotive, and aerospace industries for four years before enlisting in the Air Force. He attended Officers Training School and received a commission as a second lieutenant in May 1985. After completing Undergraduate Navigator Training and Electronic Warfare Officer Training School, Mr. Smith was assigned to the 20th Tactical Fighter Wing at RAF Upper Heyford, U.K., where he served as an F-111E Weapons Systems Officer and and an EF-111A Electronic Warfare Officer. Mr. Smith entered the the Graduate Systems Engineering Program, School of Engineering, Air Force Institute of Technology in May 1992. He separated from the Air Force in Dec 1992, and completed the AFIT degree program as a civilian.

Permament Address: 4869 Butler
Troy, Michigan 48098

Captain Brian K. Standley was born 28 August 1966 in Bridgeport, Connecticut. He graduated from Trumbull High School in Trumbull, Connecticut. He attended Worcester Polytechnic Institute, graduating with a Bachelor of Science in Mechanical Engineering in May 1988. While in college, he earned a commission as a second lieutenant through the Reserve Officers Training Corps. Captain Standley entered the U. S. Air Force and was stationed at George AFB, California. He was assigned to the 35th Equipment Maintenance Squadron and held the positions of Munitions Branch OIC and later Maintenance Branch OIC. Captain Standley entered the Graduate Systems Engineering Program, Graduate School of Engineering, Air Force Institute of Technology in May 1992.

Permament Address: 4300 Maxlin Road
Kettering, Ohio 45429

Captain Tami L. Volk was born on 29 May 1966 in Great Falls, Montana. She graduated from Charles M. Russell High School in Great Falls in June 1984. Later that fall she entered the School of Engineering at Montana State University, in Bozeman, Montana. During her sophomore year, Tami joined the ROTC program. She graduated with honors in Mechanical Engineering, a distinguished graduate from ROTC, and received a regular commission in 1988. Her first assignment was to the Ballistic Systems Division at Norton AFB in San Bernardino, California. She served as the Assembly and Checkout Support Equipment Manager for Peacekeeper Rail Garrison and later as a Systems Engineer on staff. Tami entered the School of Engineering at the Air Force Institute of Technology in May 1992 to study Systems Engineering. She married Captain Richard H. Painter Jr. in March 1993 during her stay at AFIT.

Permament Address: 755 33 Avenue NE
Great Falls, Montana 59404

Captain Linda Bilewski Wildes was born on 1 May 1961 in Albany, New York. She graduated from Medway High School in Medway, Massachusetts in 1979. After high school, she received an AFROTC scholarship at Northeastern University, Boston, Massachusetts where she graduated with honors and a Bachelor of Science degree in Chemical Engineering in June 1984. She was commissioned as an officer in the United States Air Force that same year and in August 1984, arrived at the Materiel Management Directorate, Ogden Air Logistic Center, Hill AFB, Utah for her first assignment, as Corrosion Control Program Manager. Captain Wildes was then reassigned as Software Systems Manager for ICBMs in January 1989. Later that year she received orders for reassignment to the Aero Propulsion and Power Directorate, Wright Laboratory, Wright-Patterson AFB, Ohio as Advanced Fuels Research Group Leader in September 1989. In May 1992, Captain Wildes entered the School of Engineering at the Air Force Institute of Technology.

Captain Wildes married Captain Francis Patrick Wildes of Brooklyn, New York in 1989. They have two children: Alexandra Marie and Daniel Francis.

Permament Address: 12 Sunset Road
Key Largo, Florida 31037

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1993	Master's Thesis

4. TITLE AND SUBTITLE
CREATION OF PROTOTYPE AIRCREW PROTECTION EQUIPMENT BASED ON FACE ANTHROPOMETRY

5. FUNDING NUMBERS

6. AUTHOR(S)
Martin E. Piccus, Capt, USAF
Brian K. Standley, Capt, USAF
Linda B. Wildes, Capt, USAF
Mr Gary A. Smith
Tami L. Volk, Capt, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GSE/ENY/93D-2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Human Systems Center / Life Support Systems Division (HSC/YAS)
8107 13th Street
Brooks AFB, Texas 78235-5218

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The objective of this research is to develop a process which will analyze and group human faces, concentrating on the oxygen mask seal area, into a minimum number of generic yet distinct face types. The shape and size, or form, of the face is described by bending energies. This bending energy provides the features for the grouping of similar face types, while distinguishing representatives for each cluster. The face tariffs are the percentage of faces grouped under a representative face. The representative faces are then transformed into three-dimensional solids using computer-aided design. A rapid prototyper creates the solids or plugs from which MBU-5/P masks are made and tested. The two major outcomes of this research are a formal morphological method which allows the comparison between multiple faces, and a process to transform anthropomorphic surfaces into a prototype mask mold.

14. SUBJECT TERMS
anthropometry measurement, form characterization, multivariate analysis, fuzzy sets, clustering, fuzzy clustering, rapid prototyping, shape analysis, oxygen masks, variogram, structural analysis, kriging, bending energy

15. NUMBER OF PAGES
321

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL